

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 15, 2004, 15:05:46 ; Search time 12084 Seconds  
(without alignments)  
11373.780 Million cell updates/sec

Title: US-09-721-114-1\_COPY\_655\_3825

Perfect score: 3171

Sequence: 1 atggagattgttcagtaga.....acgagtattggtaccgc 3171

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.scs.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.mam.\*

37: em.htg.vrt.\*

38: em.sv.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3171	100.0	4310	6	BD083699	BD083699 Novel gen
2	3167.8	99.9	3896	8	AF326768	AF326768 Oryza sat
3	2978.4	93.9	3626	8	AK103364	AK103364 Oryza sat
4	2573.2	81.1	9455	6	BD083700	BD083700 Novel gen
5	2573.2	81.1	150594	8	AP001859	AP001859 Oryza sat
6	62.8	2.0	7218	6	I66494	I66494 Sequence 14
7	62	2.0	2000	6	AX655393	AX655393 Sequence
8	52	1.6	2000	6	AX655393	AX655393 Sequence
9	50.8	1.6	180023	2	EX897685	EX897685 Danio rer
10	49.8	1.6	178469	2	EX248504	EX248504 Danio rer
11	48.8	1.5	185533	2	EX005453	EX005453 Danio rer
12	48.6	1.5	228730	2	AC096088	AC096088 Rattus no
13	48.4	1.5	124927	5	EX510987	EX510987 Zebrafish
14	48.4	1.5	250029	3	AE014820	AE014820 Plasmodiu
15	48.2	1.5	141041	2	AC026135	AC026135 Homo sapi
16	48.2	1.5	148130	9	AC092925	AC092925 Homo sapi
17	47.8	1.5	213544	2	EX510939	EX510939 Danio rer
18	47.2	1.5	106172	9	AL591122	AL591122 Human DNA
19	46.8	1.5	225453	2	AC097957	AC097957 Rattus no
20	46.6	1.5	127494	2	AC117775	AC117775 Mus muscu
21	46.6	1.5	187418	10	AL672278	AL672278 Mouse DNA
22	46.2	1.5	155304	9	AC055845	AC055845 Homo sapi
23	46.2	1.5	167446	2	AC137718	AC137718 Homo sapi
24	46.2	1.5	189203	2	AC026077	AC026077 Homo sapi
25	45.6	1.4	198743	2	EX530070	EX530070 Danio rer
26	45.4	1.4	274	6	AR248600	AR248600 Sequence
27	45.4	1.4	145120	8	AP003434	AP003434 Oryza sat
28	45.2	1.4	1141	6	AX083744	AX083744 Sequence
29	45.2	1.4	93100	9	AL162614	AL162614 Human DNA
30	45.2	1.4	94534	5	AL929250	AL929250 Zebrafish
31	45.2	1.4	114980	2	AC087825	AC087825 Homo sapi
32	45.2	1.4	158392	2	AC022048	AC022048 Homo sapi
33	45	1.4	173233	9	AC116002	AC116002 Homo sapi
34	45	1.4	217726	10	AL928680	AL928680 Mouse DNA
35	45	1.4	330050	3	PPA929355	PPA929355 Plasmodiu
36	44.8	1.4	139074	2	AC131191	AC131191 Mus muscu
37	44.8	1.4	165488	10	AL807810	AL807810 Mouse DNA
38	44.4	1.4	3400	14	SH10145A	L20708 Pseudorabie
39	44.4	1.4	5200	14	SHE58096S	AJ580965 Suid herp
40	44.4	1.4	151802	3	AC114263	AC114263 Dictyoste
41	44.4	1.4	246642	2	AC123079	AC123079 Rattus no
42	44.2	1.4	95097	9	AL133232	AL133232 Human DNA
43	44.2	1.4	177747	10	AC122051	AC122051 Mus muscu
44	44.2	1.4	182528	2	AC126422	AC126422 Mus muscu
45	44.2	1.4	195509	2	AC107868	AC107868 Mus muscu

# ALIGNMENTS

RESULT 1  
BD083699  
LOCUS  
DEFINITION Novel gene participating in response to brassinosteroid.  
ACCESSION BD083699  
VERSION BD083699.1 GI:22629309  
KEYWORDS JP 2001327287-A/1.  
SOURCE Oryza sativa  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Erihartoideae; Oryzeae; Oryza.  
1 (bases 1 to 4310)  
Hirochika, H., Yamazaki, T. and Miyao, A.  
REFERENCE  
AUTHORS  
TITLE Novel gene participating in response to brassinosteroid

JOURNAL Patent: JP 2001327287-A 1 27-NOV-2001;  
 NATL INST OF AGROBIOLOGICAL RESOURCES, BIO ORIENTED TECHNOLOGY  
 RESEARCH ADVA JAPAN FORAGE SEED ASSOCIATION  
 COMMENT OS Oryza sativa (rice)  
 PN JP 2001327287-A/1  
 PD 27-NOV-2001  
 PF 19-MAY-2000 JP 2000149106  
 PI HIROHIKO HIROCHIKO, TOKIO YAMAZAKI, AKIO MIYAO  
 PC C12N15/09, C12N15/00  
 CC Novel gene participating in response to brassinosteroid FH  
 Key Location/Qualifiers  
 FT CDS (655)..(3828).  
 FEATURES  
 source Location/Qualifiers  
 1..4310  
 /organism="Oryza sativa"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4530"  
 ORIGIN  
 Query Match 100.0%; Score 3171; DB 6; Length 4310;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGCGAATTGTTGCAGTAGATCAGGAGGAGCTCGTGTGTTGGACGCAACTGTATGCTT 60  
 DB 655 ATGCGAATTGTTGCAGTAGATCAGGAGGAGCTCGTGTGTTGGACGCAACTGTATGCTT 714  
 QY 61 GCTCGTGTGGAACTCGGTGCTGTAGCCGACGATGTTGGAGCTGACGACGCGCTGTCAG 120  
 DB 715 GCTCGTGTGGAACTCGGTGCTGTAGCCGACGATGTTGGAGCTGACGACGCGCTGTCAG 774  
 QY 121 GATGCAAGCCGCTGAAGCTGGTGTAGACGAACCGGCACAAACCAATGCGAGCAATTTCTCC 180  
 DB 775 GATGCAAGCCGCTGAAGCTGGTGTAGACGAACCGGCACAAACCAATGCGAGCAATTTCTCC 834  
 QY 181 ATAAGAGGGTATGTTGCTCTTCTTACAGAAAGATCCAAATTCGCTCTCTATCTCGG 240  
 DB 835 ATAAGAGGGTATGTTGCTCTTCTTACAGAAAGATCCAAATTCGCTCTCTATCTCGG 894  
 QY 241 ATTTTCATGACACAGAAAAAATGTGATGAACACAAAGCTAGTTCAAGCCCAATTTCTGTA 300  
 DB 895 ATTTTCATGACACAGAAAAAATGTGATGAACACAAAGCTAGTTCAAGCCCAATTTCTGTA 954  
 QY 301 GCAAGTTTCGACGATGGGATTTGCTGGAAGTCTTGGATGATGTAAGATTCAGATAAT 360  
 DB 955 GCAAGTTTCGACGATGGGATTTGCTGGAAGTCTTGGATGATGTAAGATTCAGATAAT 1014  
 QY 361 GGAACGACCAAGAATCTTCCGCAAAACAGATGGCAAAAGTGTATGTTGCTCCATC 420  
 DB 1015 GGAACGACCAAGAATCTTCCGCAAAACAGATGGCAAAAGTGTATGTTGCTCCATC 1074  
 QY 421 ACATTTGTCGGAGCACTTTTGGCCCTGTAGTGTGGTTCACAAAGATGTCCTGAGC 480  
 DB 1075 ACATTTGTCGGAGCACTTTTGGCCCTGTAGTGTGGTTCACAAAGATGTCCTGAGC 1134  
 QY 481 ACACAATCATCTCAAGGGAAGATGCTGTATAGATCAACTCTTCAAGAGTGTGCAAGAA 540  
 DB 1135 ACACAATCATCTCAAGGGAAGATGCTGTATAGATCAACTCTTCAAGAGTGTGCAAGAA 1194  
 QY 541 GGCATGACTCCAAATGCAATGCGCTTTCTGGCAGAAATGAGCTGCTGAGGCCAATACT 600  
 DB 1195 GGCATGACTCCAAATGCAATGCGCTTTCTGGCAGAAATGAGCTGCTGAGGCCAATACT 1254  
 QY 601 GATTACCAATGAAGAATTTGCAAGGCCAGCCCAAAATTTATGATGGCAGCAAAATGTC 660  
 DB 1255 GATTACCAATGAAGAATTTGCAAGGCCAGCCCAAAATTTATGATGGCAGCAAAATGTC 1314  
 QY 661 TCTGAGGACAAACACTTCTGTGTGATGTTGGGGCTTTACTGTAAGTTCCCGAGATTACATGG 720  
 DB 1315 TCTGAGGACAAACACTTCTGTGTGATGTTGGGGCTTTACTGTAAGTTCCCGAGATTACATGG 1374  
 QY 721 CACATAGAAGTAAATGGTGCAGATCAACCTCCATCCACTCCAAACATTTCTGAAGTGGTC 780



Query Match		99.9%	Score 3167.8	DB 8	Length 3896
Best Local Similarity		99.9%	Pred. No. 0		
Matches 3169		Conservative	0	Mismatches 2	Indels 0
Qy	1	ATGGAGATTGTTGCACTAGATCAGAGGGAGCTGCTGTTGTTGGGACGAJCTGTATGCTT	60		
Db	241	ATGGAGATTGTTGCACTAGATCAGAGGGAGCTGCTGTTGTTGGGACGAJCTGTATGCTT	300		
Qy	61	GCTCGTGTGGAACTGCTCTAGCGCAGTGTGGAGCTGACAGGAGCGCTCGTCAG	120		
Db	301	GCTCGTGTGGAACTGCTCTAGCGCAGTGTGGAGCTGACAGGAGCGCTCGTCAG	360		
Qy	121	GATGCAGCGCTGAAGCTGGTGTAGCGCAACCGGCAACCAACCAATGGGAGCATTTCTCC	180		
Db	361	GATGCAGCGCTGAAGCTGGTGTAGCGCAACCGGCAACCAACCAATGGGAGCATTTCTCC	420		
Qy	181	ATPAGAGGATGTGCTCTCTTTCAGAGAGAGGATCCAAATTTCTGCTCTCTATCTCGG	240		
Db	421	ATPAGAGGATGTGCTCTCTTTCAGAGAGAGGATCCAAATTTCTGCTCTCTATCTCGG	480		
Qy	241	ATTTTCCATGACACAGAAAAATGTGTAGAACCAAAAGCTAGTTCAAGCCCAATTTCTGTA	300		
Db	481	ATTTTCCATGACACAGAAAAATGTGTAGAACCAAAAGCTAGTTCAAGCCCAATTTCTGTA	540		
Qy	301	GCAAAGTTTCGACATGGGATGCTCGAGTCTGCTGGATAGTTGGAACCTCAGATANT	360		
Db	541	GCAAAGTTTCGACATGGGATGCTCGAGTCTGCTGGATAGTTGGAACCTCAGATANT	600		
Qy	361	GGAAACAGCACCAAGAACTCTTCCCGCAAGACAGATGGCACAAAGTGTGTTGCCATC	420		
Db	601	GGAAACAGCACCAAGAACTCTTCCCGCAAGACAGATGGCACAAAGTGTGTTGCCATC	660		
Qy	421	ACATTTGTTGGAGCACTTTTGTGCTGCTAGTGTGTTTCCAAAAGTGTCTCTAGC	480		
Db	661	ACATTTGTTGGAGCACTTTTGTGCTGCTAGTGTGTTTCCAAAAGTGTCTCTAGC	720		
Qy	481	ACCAATCATCTCAAGGAGAAATGTGTATCAATCAACTCTTCCAAAGAGTGTGCAAGAA	540		
Db	721	ACCAATCATCTCAAGGAGAAATGTGTATCAATCAACTCTTCCAAAGAGTGTGCAAGAA	780		
Qy	541	GGCAATGACTCCAAATGCAATGCGCTTCTGGCAAGATGGAGCTGTGAGGCCAATACT	600		
Db	781	GGCAATGACTCCAAATGCAATGCGCTTCTGGCAAGATGGAGCTGTGAGGCCAATACT	840		
Qy	601	GATTCACCAATGGAAGTTTGAAGGGCAGCCCAAAATTAATGATGTGCGACCAATGTC	660		
Db	841	GATTCACCAATGGAAGTTTGAAGGGCAGCCCAAAATTAATGATGTGCGACCAATGTC	900		
Qy	661	TCTGAGGACCAACACTTCTGTTGATGTTGGGGCTTTTACCTGAAAGTTCCCCAGATTCATGG	720		
Db	901	TCTGAGGACCAACACTTCTGTTGATGTTGGGGCTTTTACCTGAAAGTTCCCCAGATTCATGG	960		
Qy	721	CACATAGAGTAATGTGTCAGATCACTCCATCCACTCCAAACTTTCTCAAGTGTGTC	780		
Db	961	CACATAGAGTAATGTGTCAGATCACTCCATCCACTCCAAACTTTCTCAAGTGTGTC	1020		
Qy	781	CTCAAAAGAAATGAAGATGAAAATGGAATAACTGGAAGACTCTTTGCTGAGCAGTGC	840		
Db	1021	CTCAAAAGAAATGAAGATGAAAATGGAATAACTGGAAGACTCTTTGCTGAGCAGTGC	1080		
Qy	841	AAATTGACCAAGATTCCTAACCCTATGTCGGAAGAGACGTGATCAGGTTCTGAGCAG	900		
Db	1081	AAATTGACCAAGATTCCTAACCCTATGTCGGAAGAGACGTGATCAGGTTCTGAGCAG	1140		
Qy	901	TGCAATTTGACCAAGATTCGGAACCAAGTGTCTGGGACAGAAATGTGAGCAGATCTCGCAAT	960		
Db	1141	TGCAATTTGACCAAGATTCGGAACCAAGTGTCTGGGACAGAAATGTGAGCAGATCTCGCAAT	1200		
Qy	961	GAGCCATGTGAAGAGTTGTTCTCAAAAGAGCTCCAAATCTTAAGAGGAGAGCGGATAG	1020		
Db	1201	GAGCCATGTGAAGAGTTGTTCTCAAAAGAGCTCCAAATCTTAAGAGGAGAGCGGATAG	1260		



2341	Db	 CCACATCTTCAGAACTTTTCAGTCTACTCAAGAAACAGCAGACACATTTGCGATGGGAAGAA	2400
2161	Qy	 ATGGTCACTATTGCTGCAAGCTCACCACCTATTTTTCACATCATGATGATCAGTATATTGCT	2220
2401	Db	 ATGGTCACTATTGCTGCAAGCTCACCACCTATTTTTCACATCATGATGATCAGTATATTGCT	2460
2221	Qy	 GAAGCAACCACTGAAACATTTGGGGCGGTAAAGACGCAAGAAGCTAAACGTCGGAGCAATTT	2280
2461	Db	 GAAGCAACCACTGAAACATTTGGGGCGGTAAAGACGCAAGAAGCTAAACGTCGGAGCAATTT	2520
2281	Qy	 AAGGCCACTACAAGAAATTTCTCCAGCAGCAACCATGTGGTGTCTCAATTTAGACCTGGTATC	2340
2521	Db	 AAGGCCACTACAAGAAATTTCTCCAGCAGCAACCATGTGGTGTCTCAATTTAGACCTGGTATC	2580
2341	Qy	 CAAGCAGTTGACTTGACTTCTACTCATGTCTATGGGATCTTCCAGCAATTTATGCATCTCGC	2400
2581	Db	 CAAGCAGTTGACTTGACTTCTACTCATGTCTATGGGATCTTCCAGCAATTTATGCATCTCGC	2640
2401	Qy	 CAACCAGTAAATGGCGCCACTGGACCGGTATGTCTGAAAGAGCGGTAAACCAAGTCCATGCA	2460
2641	Db	 CAACCAGTAAATGGCGCCACTGGACCGGTATGTCTGAAAGAGCGGTAAACCAAGTCCATGCA	2700
2461	Qy	 AGAAATTTTCCAGCACAAAATAGCAACCAATGGAGCGGAGTAGTATGTGATCGGAGAAAT	2520
2701	Db	 AGAAATTTTCCAGCACAAAATAGCAACCAATGGAGCGGAGTAGTATGTGATCGGAGAAAT	2760
2521	Qy	 GCTGGCAAGCTAGTCTTGATCTCTAAGAAATCCATGCTCGAAGCGCATCTTCTGAGAAATG	2580
2761	Db	 GCTGGCAAGCTAGTCTTGATCTCTAAGAAATCCATGCTCGAAGCGCATCTTCTGAGAAATG	2820
2581	Qy	 ATGGATCCATCAACATTAGCAAGCTTCCCCAACTATGGAACTTCTAGCAGGAACCAAGATG	2640
2821	Db	 ATGGATCCATCAACATTAGCAAGCTTCCCCAACTATGGAACTTCTAGCAGGAACCAAGATG	2880
2641	Qy	 GAGTCTCAACTTTCATTAATTTCTCAGTAGCACAATAATCAGTACAAGAGATCAACCAAGACA	2700
2881	Db	 GAGTCTCAACTTTCATTAATTTCTCAGTAGCACAATAATCAGTACAAGAGATCAACCAAGACA	2940
2701	Qy	 TCATATGGCAGTAACTGAAATGSAAGATTTCCATTGACATTTGGAAGACTTATCAAGCAAT	2760
2941	Db	 TCATATGGCAGTAACTGAAATGSAAGATTTCCATTGACATTTGGAAGACTTATCAAGCAAT	3000
2761	Qy	 CAGCTGCATGATCTGCAAGACCTTTTACGCCCAATCCTAGAGTTGGTGTGCTTGGCTCC	2820
3001	Db	 CAGCTGCATGATCTGCAAGACCTTTTACGCCCAATCCTAGAGTTGGTGTGCTTGGCTCC	3060
2821	Qy	 TTGCTGCAGAAGGAAAATTCGAAAATGGTGTGGAGAACTGTGGCACAACAATCTGGTTATAAG	2880
3061	Db	 TTGCTGCAGAAGGAAAATTCGAAAATGGTGTGGAGAACTGTGGCACAACAATCTGGTTATAAG	3120
2881	Qy	 TTAGGAGTGTCAACAGGAAATACATCCGATCAGATGACAGAAAGGAGACATTTTGAGCC	2940
3121	Db	 TTAGGAGTGTCAACAGGAAATACATCCGATCAGATGACAGAAAGGAGACATTTTGAGCC	3180
2941	Qy	 CTGAATTTCTGGAATGTTTTTCAGCAAAATGGAATGCAATTCAGATTTGGTCTTGTAGTCTCC	3000
3181	Db	 CTGAATTTCTGGAATGTTTTTCAGCAAAATGGAATGCAATTCAGATTTGGTCTTGTAGTCTCC	3240
3001	Qy	 AGTCGAGATTTTTCATCCGAGGAAACAGCATAGCTCATCTTTCGACCAAGGAGGCAAGGGT	3060
3241	Db	 AGTCGAGATTTTTCATCCGAGGAAACAGCATAGCTCATCTTTCGACCAAGGAGGCAAGGGT	3300
3061	Qy	 AAAATGGTTTCATCCCTTGATCGGTTTGTGTGAGACAGGATATCTGTATATACTAAACAAGAC	3120
3301	Db	 AAAATGGTTTCATCCCTTGATCGGTTTGTGTGAGACAGGATATCTGTATATACTAAACAAGAC	3360
3121	Qy	 CCAGCTGATTTTTCATCAAAATCAGTAAACGATTAACGAGTATATGGMTACCCG	3171
3361	Db	 CCAGCTGATTTTTCATCAAAATCAGTAAACGATTAACGAGTATATGGMTACCCG	3411

Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Nariawa, R., Niihara, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.  
Genome Exploration Laboratory Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, K., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Ito, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ono, M., Osato, N., Ota, Y., Saiboh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Soabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.  
Location/Qualifiers  
1. 3626  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="rRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="J033126N23"

## ORIGIN

Query Match 93.9%; Score 2978.4; DB 8; Length 3626;  
Best Local Similarity 100.08; Pred. No. 0;  
Matches 2979; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGAGATGTTGCAGTAGATCAGGAGGAGCTCGTGTGTTGGGAGCAACTGTATGCTT 60  
DB 647 ATGGAGATGTTGCAGTAGATCAGGAGGAGCTCGTGTGTTGGGAGCAACTGTATGCTT 706

QY 61 GCTCGTGTGGAACCTGGTGTAGCGCCAGTGTGGAGCTGACAGGAGCGCTCGTFCAG 120  
DB 707 GCTCGTGTGGAACCTGGTGTAGCGCCAGTGTGGAGCTGACAGGAGCGCTCGTFCAG 766

QY 121 GATGAGCGCTGTAAGCTGGTGTAGAGCAACCGGAGCAACCAATGCGAGCAATTTCTCC 180  
DB 767 GATGAGCGCTGTAAGCTGGTGTAGAGCAACCGGAGCAACCAATGCGAGCAATTTCTCC 826

QY 181 ATAGAGGGATGTTGCTCTTTCAGAGAGAGATCCAAATTCCTCTCTATCTCGG 240  
DB 827 ATAGAGGGATGTTGCTCTTTCAGAGAGAGATCCAAATTCCTCTCTATCTCGG 886

QY 241 ATTTTCCATGCCAGAAAATGATGAACACAAAGCTAGTTCAAGCCCAATTTCTGTA 300  
DB 887 ATTTTCCATGCCAGAAAATGATGAACACAAAGCTAGTTCAAGCCCAATTTCTGTA 946

QY 301 GCAGAGTTTCAGCAGTGGATGCTCGAGGTGCTTGGATAGTTGAAAGCTCAGATAAT 360  
DB 947 GCAGAGTTTCAGCAGTGGATGCTCGAGGTGCTTGGATAGTTGAAAGCTCAGATAAT 1006

QY 361 GGAACAGCACCAAGAACTTTCCCGCAAGCAGAAATGGCAAGTGTGTTCTCCATC 420  
DB 1007 GGAACAGCACCAAGAACTTTCCCGCAAGCAGAAATGGCAAGTGTGTTCTCCATC 1066

QY 421 ACATTTGTTGGAGCACTTTTGCTGCTAGTGTGTTGCCAAAGAGTGTCTCTAGC 480  
DB 1067 ACATTTGTTGGAGCACTTTTGCTGCTAGTGTGTTGCCAAAGAGTGTCTCTAGC 1126

QY 481 ACACAAATCATCTCAAGGGAAGATGCTGATAGATCAACTCTTCCAAAGAGTGTGGAAGA 540  
DB 1127 ACACAAATCATCTCAAGGGAAGATGCTGATAGATCAACTCTTCCAAAGAGTGTGGAAGA 1186

QY 541 GGCATGACTCCAAATGCAATGCGCTTCTGGCAAGAAATGGAGTGTGTTGAGGCGCAATCT 600  
DB 1187 GGCATGACTCCAAATGCAATGCGCTTCTGGCAAGAAATGGAGTGTGTTGAGGCGCAATCT 1246

QY 601 GATTCCCAATGAAGATTTCGAGGCGCAGCCCAAAATTTATGATGTGCGCAGCAATGTC 660  
DB 1247 GATTCCCAATGAAGATTTCGAGGCGCAGCCCAAAATTTATGATGTGCGCAGCAATGTC 1306

QY 661 TCTGAGGACACACACTTCTGTTGATGTTGGGCTTTTACTCTGAAGTTTCCCGAGATTACATGG 720  
DB 1307 TCTGAGGACACACACTTCTGTTGATGTTGGGCTTTTACTCTGAAGTTTCCCGAGATTACATGG 1366

QY 721 CACATAGAAGTAATGTTGTCAGATCAACTCTCCATCTCCATCTCCAAAGCTTTCTGAAGTGTGTC 780  
DB 1367 CACATAGAAGTAATGTTGTCAGATCAACTCTCCATCTCCAAAGCTTTCTGAAGTGTGTC 1426

QY 781 CTCAAAGAAATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 840  
DB 1427 CTCAAAGAAATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 1486

QY 841 AATTTGACCAAGATTCCTAAACCCAAAGTCTGGAAGGAACTGATGATCAGGTTGTCGAGCAG 900  
DB 1487 AATTTGACCAAGATTCCTAAACCCAAAGTCTGGAAGGAACTGATGATCAGGTTGTCGAGCAG 1546

QY 901 TGCATTTGACCAAGATTCCTAAACCCAAAGTCTGGAAGGAACTGATGATCAGGATCTGCAAT 960  
DB 1547 TGCATTTGACCAAGATTCCTAAACCCAAAGTCTGGAAGGAACTGATGATCAGGATCTGCAAT 1606

QY 961 GAGCCATGTAAGAGTGTGTTCTCAAAGAGCTCCAAATCTAAGAGGAAAGACCGGATAAG 1020  
DB 1607 GAGCCATGTAAGAGTGTGTTCTCAAAGAGCTCCAAATCTAAGAGGAAAGACCGGATAAG 1666

QY 1021 AAGTTGATGAAGAGCAGCAGCAGCAGCAAGAAAGCCACTGCCAGGCTGATGTTTCAGAT 1080  
DB 1667 AAGTTGATGAAGAGCAGCAGCAGCAGCAAGAAAGCCACTGCCAGGCTGATGTTTCAGAT 1726

QY 1081 GCAAGCTTTTCGAGAAAGCCAAAGAGTGGGCTTCTATCAGAAATTTATTAATGCT 1140  
DB 1727 GCAAGCTTTTCGAGAAAGCCAAAGAGTGGGCTTCTATCAGAAATTTATTAATGCT 1786

QY 1141 AACAGGTTGAGGATTTCTAGAGTGAAGAGTTCATGTTGAAATGCGGCTGATCCCTGT 1200  
DB 1787 AACAGGTTGAGGATTTCTAGAGTGAAGAGTTCATGTTGAAATGCGGCTGATCCCTGT 1846

QY 1201 GAGGATGATAGAGTACCATCCCGTCCGATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1260  
DB 1847 GAGGATGATAGAGTACCATCCCGTCCGATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1906

QY 1261 AACCATACAGTGGGAGAGATGGGTTAAATCAAGTAAGAACCAAGACAAACCGAAATAC 1320  
DB 1907 AACCATACAGTGGGAGAGATGGGTTAAATCAAGTAAGAACCAAGACAAACCGAAATAC 1966

QY 1321 TCTGATGTTGATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
DB 1967 TCTGATGTTGATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2026

QY 1381 ACTGGAAGTGTGCATCACACAGTTGCTCATCCAGCTGGGAATTTGACCAACAAAGATG 1440  
DB 2027 ACTGGAAGTGTGCATCACACAGTTGCTCATCCAGCTGGGAATTTGACCAACAAAGATG 2086

QY 1441 ACACCCACTGCGATCTCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500  
DB 2087 ACACCCACTGCGATCTCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2146

QY 1501 AATATGATAGACAGATGCTGTCAGCAGTGTATCAGAAATCTCCACACAGAGTGTCTCA 1560  
DB 2147 AATATGATAGACAGATGCTGTCAGCAGTGTATCAGAAATCTCCACACAGAGTGTCTCA 2206

QY 1561 TCAGAGGGGAAAACAGCGGGTTTGAAGGAGGAAAACACATTCNGCTGTGTAGTACCACA 1620  
DB 2207 TCAGAGGGGAAAACAGCGGGTTTGAAGGAGGAAAACACATTCNGCTGTGTAGTACCACA 2266

QY 1621 TATGTTGTTGAAGACACACAGAAATGTTGAGAAATATATATGATCTCAGCGCAGAGATCAA 1680  
DB 2267 TATGTTGTTGAAGACACACAGAAATGTTGAGAAATATATATGATCTCAGCGCAGAGATCAA 2326

QY 1681 TCCAGATGGAACCGGAAAACCTGTTTCTGAGTCACTGCGCAAGGTTTCTCCAGCTGAG 1740  
DB 2327 TCCAGATGGAACCGGAAAACCTGTTTCTGAGTCACTGCGCAAGGTTTCTCCAGCTGAG 2386

QY 1741 CATGATATCCAAATATATGCTGACCTTCATGACGAGAGTCTACCCAGAGAAAGAAAG 1800  
Db CATGATATCCAAATATATGCTGACCTTCATGACGAGAGTCTACCCAGAGAAAGAAAG 2446  
QY 1801 CAAAACTTGAAGTACTCGTGAAGAAACAGACCATGATAGATGACATCCCATCGGATATT 1860  
Db CAAAACTTGAAGTACTCGTGAAGAAACAGACCATGATAGATGACATCCCATCGGATATT 2506  
QY 1861 GTTGAACCTGTAGTAAACACAGCATGAGAGGAGCTTATGACTGAGACTGATTTCT 1920  
Db GTTGAACCTGTAGTAAACACAGCATGAGAGGAGCTTATGACTGAGACTGATTTCT 2566  
QY 1921 GACATCAACCTGATTCATCAACAGCAACCTGCTCATGATGATTTGTAAATAGTGTGCC 1980  
Db GACATCAACCTGATTCATCAACAGCAACCTGCTCATGATGATTTGTAAATAGTGTGCC 2626  
QY 1981 AAGGATGGTTCAGATATGCAATCAAGTGTGTTGACACATAATTCCTCCACAGAGTCTTG 2040  
Db AAGGATGGTTCAGATATGCAATCAAGTGTGTTGACACATAATTCCTCCACAGAGTCTTG 2686  
QY 2041 GCATCCCAAGTACACAGAGGAGTTACAGGTCATTTGGCATTGACACACAGAGTCT 2100  
Db GCATCCCAAGTACACAGAGGAGTTACAGGTCATTTGGCATTGACACACAGAGTCT 2746  
QY 2101 CCATATCTCTCAGAACTTTCACTCTAAGGACAGCAGACACATTTCCGGATGGAAGAA 2160  
Db CCATATCTCTCAGAACTTTCACTCTAAGGACAGCAGACACATTTCCGGATGGAAGAA 2806  
QY 2161 ATGTCATATGCTGACGCTCACCATTATTTTCACATCATGATGATCAGTATATTGCT 2220  
Db ATGTCATATGCTGACGCTCACCATTATTTTCACATCATGATGATCAGTATATTGCT 2866  
QY 2221 GAAGCACCACCTGAACATTTGGGGCGCTAAGGACGCAAGAGCTAAACGTTGGAGCAATTT 2280  
Db GAAGCACCACCTGAACATTTGGGGCGCTAAGGACGCAAGAGCTAAACGTTGGAGCAATTT 2926  
QY 2281 AAGGCCATCAAGAAATTTCCAGCAGCAACATGCTGTCTCAATTTAGACCTGGTATC 2340  
Db AAGGCCATCAAGAAATTTCCAGCAGCAACATGCTGTCTCAATTTAGACCTGGTATC 2986  
QY 2341 CAAGCAGTTGACTTGTACTTCTACTCATGCTGATGCTTCCAGCAATTTATGCTCTCGC 2400  
Db CAAGCAGTTGACTTGTACTTCTACTCATGCTGATGCTTCCAGCAATTTATGCTCTCGC 3046  
QY 2401 CAACAGTAATTTGGCCCACTGGACCGCTATGCTGAAAGAGCGGTTAACCGGTCCATGCA 2460  
Db CAACAGTAATTTGGCCCACTGGACCGCTATGCTGAAAGAGCGGTTAACCGGTCCATGCA 3106  
QY 2461 AGAAATTTTCCAGCACAATAGCACCATGGAAGCGAGTATGATGATGATGATGATGATG 2520  
Db AGAAATTTTCCAGCACAATAGCACCATGGAAGCGAGTATGATGATGATGATGATGATGATG 3166  
QY 2521 GCTGGACAAGTAGTCTTGTATCTTAAAGATCCATGCTGGAGCGCATCTTCTGAGAAATG 2580  
Db GCTGGACAAGTAGTCTTGTATCTTAAAGATCCATGCTGGAGCGCATCTTCTGAGAAATG 3226  
QY 2581 ATGGATTCATCAACATTTAGCAAGCTTCCCACTATGGAATCTTCCAGCAGCAACAGATG 2640  
Db ATGGATTCATCAACATTTAGCAAGCTTCCCACTATGGAATCTTCCAGCAGCAACAGATG 3286  
QY 2641 GAGTCTCAACTTCATAATTTCTCAGTATGACACATAATCAGTACAAAGGATCAACACAGCA 2700  
Db GAGTCTCAACTTCATAATTTCTCAGTATGACACATAATCAGTACAAAGGATCAACACAGCA 3346  
QY 2701 TCATATGGCAGTAACCTGATGGAAGATTCATTTGACATTTGGAAGATTTATCAGGAT 2760  
Db TCATATGGCAGTAACCTGATGGAAGATTCATTTGACATTTGGAAGATTTATCAGGAT 3406  
QY 2761 CAGCTGCAATGCTGACAGACTTTTACGCCCAATCCTAGAGTTGGTGTGCTTGGCTCC 2820  
Db CAGCTGCAATGCTGACAGACTTTTACGCCCAATCCTAGAGTTGGTGTGCTTGGCTCC 3466

QY 2821 TTGCTGAGAGAGAAATTCGAACTGGTCGGAGAACTGTGGCAACAATCTGTGTTATAAG 2880  
Db TTGCTGAGAGAGAAATTCGAACTGGTCGGAGAACTGTGGCAACAATCTGTGTTATAAG 3526  
QY 2881 TTAGGAGTGTCAACAGGATATACATCGCATCGATGACAGAGAAAGCAATTTGAGGCC 2940  
Db TTAGGAGTGTCAACAGGATATACATCGCATCGATGACAGAGAAAGCAATTTGAGGCC 3586  
QY 2941 CTGAATTTCTGGATGTTTTCAGCAAAATGGAATGCATTGC 2980  
Db CTGAATTTCTGGATGTTTTCAGCAAAATGGAATGCATTGC 3626  
  
RESULT 4  
BD083700  
LOCUS BD083700 9455 bp DNA linear PAT 27-AUG-2002  
DEFINITION Novel gene participating in response to brassinosteroid.  
ACCESSION BD083700  
VERSION BD083700.1 GI:22629310  
KEYWORDS JP 2001327287-A/2.  
SOURCE Oryza sativa  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
1 (bases 1 to 9455)  
REFERENCE  
AUTHORS Hirochika, H., Yamazaki, T. and Miyao, A.  
TITLE Novel gene participating in response to brassinosteroid  
JOURNAL Patent: JP 2001327287-A 2 27-NOV-2001;  
NATL INST OF AGRICULTURAL RESOURCES, BIO ORIENTED TECHNOLOGY  
RESEARCH ADVANCEMENT INSTITUTION  
COMMENT  
OS Oryza sativa (rice)  
PN JP 2001327287-A/2  
PD 27-NOV-2001  
PF 19-MAY-2000 JP 2000149106  
PI HIROHICO HIROCHIKA, TOKIO YAMAZAKI, AKIO MIYAO  
PC C12N15/09, C12N15/00  
CC Novel gene participating in response to brassinosteroid PH  
Key Location/Qualifiers  
FT source 1.9455  
/organism="Oryza sativa (rice)"  
/location="Oryza sativa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4530"  
  
FEATURES  
source  
1.9455  
/organism="Oryza sativa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4530"  
  
ORIGIN  
Query Match 81.1%; Score 2573.2; DB 6; Length 9455;  
Best Local Similarity 87.8%; Pred. No. 0;  
Matches 3003; Conservative 0; Mismatches 23; Indels 396; Gaps 2;  
QY 146 ACGNACCGGCACACACCAATCGAGCATTTCCATAAGAGGATGTTGCTCTTCTTC 205  
Db ATGCATCTTTATATATGATATCTTATTTTCTGTACAGAGGATGTTGCTCTTCTTC 5506  
QY 206 AGAAGAGAGGATCCAAAATTCGCTCTCTATCTCGGATTTTCATGACAGAAAAATGTG 265  
Db AGAAGAGAGGATCCAAAATTCGCTCTCTATCTCGGATTTTCATGACAGAAAAATGTG 5566  
QY 266 ATGNACACAAAGCTAGTTCAGCCCATTTCTGTAGCAAAAGTTTCGACGATGGCATGCT 325  
Db ATGAACCAAAAGCTAGTTCAGCCCATTTCTGTAGCAAAAGTTTCGACGATGGCATGCT 5626  
QY 326 CGAAGTGTCTGGATAAGTTGAAAACCTTCAGATATGGAACAGCAGCAAGAACTCTTCCCG 385  
Db CGAAGTGTCTGGATAAGTTGAAAACCTTCAGATATGGAACAGCAGCAAGAACTCTTCCCG 5686  
QY 386 CAAGAGAGATGGCAAGTGTGTTCCATCACAATTTGTCGGGCACTTTTGTGC 445  
Db CAAGAGAGATGGCAAGTGTGTTCCATCACAATTTGTCGGGCACTTTTGTGC 5746  
QY 446 CTGCTAGTGTGGTTTCCAAAAAGTGTCTCTCTAGCACACATCATCTCAAGGAGAAATG 505





CDS  
 join(10735...10850,10933...10987,11596...11653,12285...12379,  
 12471...12574,12958...13086,13493...15302)  
 /note="unnamed protein product; Similar to Arabidopsis  
 thaliana chromosome 4, BAC clone F4D11; putative  
 myb-protein. (AL022537)"  
 /codon\_start=1  
 /protein\_id="BA094769.1"  
 /db\_xref="GI:7630236"  
 /translation="MTSDNGKADPKGEGGPPSPAFQGEISNEPKRRPLNGRTTGP  
 TRSTKGTPEEDAIISRAVOTYNGKMKIAECFFDRTDVCILRWQKVAENFELVK  
 GPKWKEDEITVOMVKLGRQKQSTIAQALPGRIGKQCRERYNHLNPGINEAMTQE  
 BEITLIHAHMYKNGMAELTKPLPGRDINSIKHNSVKKVKNYSGLHTTVSCCL  
 PLNYSACNKSPPAMTQSEDSQFAVREVSQCSOSLAKVSCSGDITTVPLG  
 CDLVANAFENKAEHDSQSMGPOACTSRAVASALPAVCHVYSSNLDPPQHLQED  
 PAQNLNLTIDENPTVPSFADNQTVCSEIENHRSLEPTDVAWEVPLSMSSQSGAEQ  
 KLFMSADPNSCLSELQWQDLSLQGLSGPDVAEISIRNSHQSDVYSSEADTH  
 FLAPYPTQTSNSSVMGLADDSPQSPVPSLCSNMTDDAPFNPRGRKEMPLSQ  
 AEVYPTQSSSGDAEMFNGCNDHRVPSSTWESIPCCDDQVNAEPFASLKEEP  
 SLTQSVTAPQDQKAGLFFYPRFPSPDLVPFVSCDLVSGDLOEFSPGILMHSITM  
 NVCTPMLWGSPTDESGTGLLSAAKSPFCTSLKGRHDLISLPDKRTEKRYGT  
 EKRGVSDTSSTGTSICINATKDDALITVLRIERSASSKSLKLVFSDENKILG  
 YTTQTKDQSGQNDHMDROTNGERSSATNATNDDLSGLVSTSSP"  
 complement(join(18608...18823,19905...20031,20072...20220,  
 20774...21113,21407...21738,21975...22508))  
 /note="unnamed protein product; ESTs  
 AU082761(S5084),D42006(S5084) correspond to a region of  
 the predicted gene.  
 Similar to Caenorhabditis elegans chromosome 3, cosmid  
 C17G10; alcohol dehydrogenase/ribitol dehydrogenase.  
 (U29739)"  
 /codon\_start=1  
 /protein\_id="BAA94770.1"  
 /db\_xref="GI:7630237"  
 /translation="MAAAAGGGGGGGGAYDPSVYDPSVKTVAHYRHYRDKNYE  
 IHOMVEGFORLSLPRDPAWPAABAVSPYCDGHDVFLLYRELRYHAYARASSSS  
 SSSPAITAGORASWANYCDLFSVLHGVNMOLPNOMLMDWDEYVYQFCQVRA  
 KLNKSDDELSOLKOPKANNVGLNYLQALVEKMSAQILEREKEGLEQFTADGY  
 DYQGSNVLKMGYSNIGLLRLHCLLDYRTGLKCLAPLIDNQOQVTVIGSHISA  
 LYHGPNLMRYAEALREFNKILLYLKQYHOKSPOYDILKKNQMAFLAVC  
 LSLCPQNLNDENVLTQLEKNDKMTKQRFDEETAAAYDELFTACPKFTIPSPFA  
 LDQPLTNYNDALQKLEFLYEVKQQLSLGYSYKLYSTITITIAKLAQYMEVEPAT  
 LRANKHSQVLYRSLTYSILMKHMHVNNKGIYSSADFDYFKIECSSIRTKPFSKN  
 LAHTIHQVGYRKSLSYFSFYFHHGINSYENGIEYRKNRKNESPAQIR"  
 join(23045...23145,24592...24670,25118...25216,25302...25415,  
 25608...25750,26364...26445,27073...27158,27527...27760,  
 29332...29917)  
 /note="unnamed protein product; Similar to Arabidopsis  
 thaliana chromosome 2 BAC T19C21; unknown protein.  
 (AC004683)"  
 /codon\_start=1  
 /protein\_id="BAA94771.1"  
 /db\_xref="GI:7630238"  
 /translation="MAGSVMDRATSDMLIGPDWAKNMEICDINRDPQSGKDVYKALK  
 KRIGHKPKYQILLALTLETIAIKNGDIIPMHVAERDVLHEMWKIVKKSQDQVKEKY  
 LTMIDTQEARGPARYQYVAYADHIVRAGAAFPKRSRDPAPLFGNSQSPAGNWS  
 PQDQREBSAGNDPALSMSIQNARGIMDLAEMNALDPGNREGLRQVYVIBLWD  
 CQRTYQRVLLVNATADELSQGLANDLQRLVAKHDAIANGAVRVEKPKSLQ  
 ALVDTDSMAQSKKEQALVDIEDPTTQETNKPSQSAVSQFPQGLALPAPFNGS  
 ATPAKSDGLVLLSMDNPSSTENSLALVPDVPVDSPTSONALVIDIIFSONMT  
 NNTAKPADPFGVNSSTLPGSQPYNAPTQPLQAOQPOQVGLYPNGGAVNPGTSYPT  
 SSGWNGQIANNATPPAQITVNYGIATVLYIYISQPFSTYIPIIGQCNHACFLCSV"  
 30166...30675  
 /note="unnamed protein product; ESTs  
 AU031545(B61825),AU031546(B61825) correspond to a region  
 of the predicted gene.  
 hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAA94772.1"  
 /db\_xref="GI:7630239"  
 /translation="MYNNQOPGAMQSPQAVTOMQSGFANQSGMPPHMPGMOPEGM  
 OPSWPQAGQPMYAOQPMWPGQFAAMPQPMYQPMQSOYRLVQOQRAQYNSQGR  
 PITYAGNMDLSQRYGLSMQDSSYMGNNSSPYSTTFSSSSSQGQPSKPKDLFGDLS  
 IAKTKQKNA"

CDS

complement(join(34629...34795,36831...36867,38803...38895,  
 41780...41830))  
 /note="unnamed protein product; hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAA94773.1"  
 /db\_xref="GI:7630240"  
 /translation="MTVVRCTTQRLLLEGRIQCDIRGKILSGKPPFSQNRKVTGI  
 EIVGLLDACVIVITVEKPENERIKKHGRLGLSLCDRLGPKUWVDSPETKVTIFY  
 VQFPRGHTLPHRG"  
 join(46367...46423,47537...47769,48031...48153,50082...50299,  
 51762...51864,52091...52204,52433...52532,52784...53004,  
 54084...54514,54601...55709,56020...57468)  
 /note="unnamed protein product; hypothetical protein"  
 /codon\_start=1  
 /protein\_id="BAA94774.1"  
 /db\_xref="GI:7630241"

CDS

/translation="MPLNGILKRLREBLDKLIAHATPPHPHPQIPVGAATAAIDIA  
 WICAPVAAAIICAPASVPDGVSPCLRAVPPPNRRTGTPSVAPSPIDIRHIFLL  
 IGFMEKDLIFMTHGVRKADDEEVGRSGSARTELKIRGLVGLMVPDGPPIHVGVDG  
 FMCVRLHVMGNSGRLALLTCSLTKANCLAMKLFPEKQOTFELGNIIEAASG  
 ERSTARVDRAPPLRRGCRKPKSPCAPLSGLELLASMTKALIPHVAEEKXIPSF  
 AGILAAFPWNLFLAILIPARISGALIHLELIEVDIYCACLVDVLLVIMDEIVAV  
 DOEGARVYGNCLMARGCTGAVPVELELTATPRQDAAANGVDEPQHQCEHSIRGY  
 VALLOKDPKPFSLSRIFHDQKKDBHKASSSPFSVAKFARMDCSKLCKITSDNGT  
 APTLPAKQNGTSDGCSITFVSTFVASVQSPQSPQSPQSPQSPQSPQSPQSPQSPQ  
 GNSKMAPSGKGAABANTDSPKDLQCPAQNDVAANVEDNTSDVDMGSLPEVQI  
 TMHIEVNGADQPPSTPKLSEVLNRNEDNGKTEETLVAQCNTLXDPNMGSGKRDQ  
 VAEQCNLTQDKPVSGQKQBOICNEPCEEVLRSSSKSKRTDKLMLKQKQSKKRTA  
 QADVAKLCKRRKPKVRLLESEINAMVEDSRSDVRHENAADPCEDDRSTIIPVME  
 VSMIDIPVSNHTVGEDEGLSKNKTKRYSDVDVDSGLSLMNLMLGKKRKTGSSVHTVAH  
 PAGNLNKVPTTASTOHDDNDTENGLELATDNMHTVCOHVSISTORCSSKGTAGL  
 SKKTSASATKIGGESTRNGONIHLVLSADQCOMETENSVLSSAKVSPAEDIQIM  
 SDHQSLDPKKKKQKLEVTREKQOTWIDIDIPMDIVELLAKNQHRLMETDSCDINR  
 IQKTEADDDPVAAQKQSDYASSVPDNTNOSKSLASQSTOKELQGLALITQESPH  
 POFQSTQEQOHLRMEEMWYIAASSPLFSDHDDQYIAEAPTEHGRKDAKLITWQF  
 KATTPRSQATCGAEPGIAQVADLTSTVMGSSSYASQPVIAPLDRAEAVQV  
 HANRFPETIATMEASLCKDRNAGQVLYPKESMPATHLLRMDPSTLASPPNYGTS  
 RNOMESOLNSOYAHKQYKSTSTSGSNLNGKIPLTFEDLSHQLDLHLRPLRPHR  
 VGLTGLLOHEIANWSENGTQSGYKLGSTGTSTHOMRKEHFEALNSGMSAKNA  
 IQGVSSSSADFLSARNSIAQSWTRGKNVHLDFRVQDICIITNKNPADFTIISND  
 NBYMDYK"  
 complement(join(58634...58667,58796...58902,59018...59083,  
 59368...59432,59585...59648,60333...60442,60620...60707,  
 60835...60945))  
 /note="unnamed protein product; ESTs  
 AU082698(R0665),D23963(R0665) correspond to a region of  
 the predicted gene  
 Similar to Arabidopsis thaliana rac GTP binding protein  
 Arac7. (AF079484)"

CDS

Query Match 81.1%; Score 2573.2; DB 8; Length 150594;  
 Best Local Similarity 87.8%; Pred. No. 0;  
 Matches 3003; Conservative 0; Mismatches 23; Indels 396; Gaps 2;  
 QY 146 ACGAAGCGGCACACCAATGCGAGCATTTCTCCATAGAGGGGTATGTTGCTCTTCTTC 205  
 DB 54044 ATGCATCTTTATATGATAATCTATATTTCTTGTCACGAGGGGTATGTTGCTCTTCTTC 54103  
 QY 206 AGAAGAGGATCCAAAATTCGTCTCTATCTCGGATTTTCCATGACCAAGAAAATGTG 265  
 DB 54104 AGAAGAGGATCCAAAATTCGTCTCTATCTCGGATTTTCCATGACCAAGAAAATGTG 54163  
 QY 266 ATGACACCAAGCTAGTTCAGCCCATTTCTCTAGCAAGTTTCCGACGATGSGATTGCT 325  
 DB 54164 ATGACCAAGCTAGTTCAGCCCATTTCTCTAGCAAGTTTTCGACGATGSGATTGCT 54223  
 QY 326 CGAAGTCTTGGATAGTTGAAAACCTTCAGATAATGGAACAGACCAAGAACTCTTCCCG 385  
 DB 54224 CGAAGTCTTGGATAGTTGAAAACCTTCAGATAATGGAACAGACCAAGAACTCTTCCCG 54283  
 QY 386 CAAGCAGAGATGCGACAAAGTGGTGTGTCCTCATCATTTGTTCCGAGCACTTTTGTGC 445  
 DB 54284 CAAGCAGAGATGCGACAAAGTGGTGTGTCCTCATCATTTGTTCCGAGCACTTTTGTGC 54343



446 CTGCTAGTGTGGTTCCTCCAAAGAGTGTCTCTAGCACACATCATCTCAAGGGAGAAATG 505  
Db |||||  
54344 CTGCTAGTGTGGTTCCTCCAAAGAGTGTCTCTAGCACACATCATCTCAAGGGAGAAATG 54403  
Qy |||||  
506 CTGATAGATCAACTCTTCCAAAGAGTGTGCAAGAGCAATGATCTCCAAATGCAATGGC 565  
Db |||||  
54404 CTGATAGATCAACTCTTCCAAAGAGTGTGCAAGAGCAATGATCTCCAAATGCAATGGC 54463  
Qy |||||  
566 CTTCTGGCAGAAATGGAGCTCTCAGGCCATATCTGATTCACCAATGAA----- 614  
Db |||||  
54464 CTTCTGGCAGAAATGGAGCTCTCAGGCCATATCTGATTCACCAATGAAAGGTATGGTAG 54523  
Qy |||||  
615 ----- 614  
Db |||||  
54524 ATGTAGAGCTTTCAAATTCCTAAGTAGAGTATTTATTTAAGGTATAGAATAAACTAATGT 54583  
Qy |||||  
615 ----- AGATTTGCAAGGCGCAGCCCAAAATTTATGATGTGGCAGCAATGT 659  
Db |||||  
54584 TTGTGTGATTTTCTCAGATTTGCAAGGCGCAGCCCAAAATTTATGATGTGGCAGCAATGT 54643  
Qy |||||  
660 CTCTGAGGACCAACTCTCTGTGTGATGTGGGCTTTACTCTGAAGTTCGCCAGATTACATG 719  
Db |||||  
54644 CTCTGAGGACCAACTCTCTGTGTGATGTGGGCTTTACTCTGAAGTTCGCCAGATTACATG 54703  
Qy |||||  
720 GCACATAGAAATGATGTGAGATCAACCTCCATCCATCCAAAATTCTTCTGAGAGTGGT 779  
Db |||||  
54704 GCACATAGAAATGATGTGAGATCAACCTCCATCCATCCAAAATTCTTCTGAGAGTGGT 54763  
Qy |||||  
780 CCTCAGAAAGAAATGAGATGAAATGGAAGAACTGAAGAGACTCTTTGTCTGAGAGCAGTG 839  
Db |||||  
54764 CCTCAGAAAGAAATGAGATGAAATGGAAGAACTGAAGAGACTCTTTGTCTGAGAGCAGTG 54823  
Qy |||||  
840 CAATTTGACCAAGATCCTTAACCCAAATGTCTGGAAGGAAAGCTGATCAGGTGTCTGAGCA 899  
Db |||||  
54824 CAATTTGACCAAGATCCTTAACCCAAATGTCTGGAAGGAAAGCTGATCAGGTGTCTGAGCA 54883  
Qy |||||  
900 GTGCATTTGACCAAGATCCTTAACCCAAATGTCTGGAAGGAAAGCTGATCAGGTGTCTGAGCA 959  
Db |||||  
54884 GTGCATTTGACCAAGATCCTTAACCCAAATGTCTGGAAGGAAAGCTGATCAGGTGTCTGAGCA 54943  
Qy |||||  
960 TGAGCCATGTGAGAGAGTGTCTCCTCAAGAGAGCTTCCAAATCTTAAGAGAGAGAGCGATAA 1019  
Db |||||  
54944 TGAGCCATGTGAGAGAGTGTCTCCTCAAGAGAGCTTCCAAATCTTAAGAGAGAGAGCGATAA 55003  
Qy |||||  
1020 GAAGTTGATGAAG 1079  
Db |||||  
55004 GAAGTTGATGAAG 55063  
Qy |||||  
1080 TGCAAGAGCTTTGTCGAG 1139  
Db |||||  
55064 TGCAAGAGCTTTGTCGAG 55123  
Qy |||||  
1140 TAACCAAGTTGAGGATTTAG 1199  
Db |||||  
55124 TAACCAAGTTGAGGATTTAG 55183  
Qy |||||  
1200 TGAGAGATGAGAGATCAACATCCCGTCCCGATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1259  
Db |||||  
55184 TGAGAGATGAGAGATCAACATCCCGTCCCGATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 55243  
Qy |||||  
1260 CAACCATACAGTGGGAG 1319  
Db |||||  
55244 CAACCATACAGTGGGAG 55303  
Qy |||||  
1320 CTTCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1379  
Db |||||  
55304 CTTCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 55363  
Qy |||||  
1380 AACTGGAAGTGTGCATCACACAGTGTCTCATCCAGCTGGGAATTTTGCAGCAACAAAAAGT 1439  
Db |||||  
55364 AACTGGAAGTGTGCATCACACAGTGTCTCATCCAGCTGGGAATTTTGCAGCAACAAAAAGT 55423

1440 GACACCCACTGCGAGTACTCAGCATGATGATGAGATGATGATGATGATGATGATGATGATGAT 1499  
Db |||||  
55424 GACACCCACTGCGAGTACTCAGCATGATGATGAGATGATGATGATGATGATGATGATGATGAT 55483  
Qy |||||  
1500 AATATGCTAATGACAGATGCTCTGTCAGCATGTATCAGAAATCTCCACACAGAGGTGCTC 1559  
Db |||||  
55484 AATATGCTAATGACAGATGCTCTGTCAGCATGTATCAGAAATCTCCACACAGAGGTGCTC 55543  
Qy |||||  
1560 ATCAAAGGGGAAACACAGCGGTTTGAATAGGGGAAACACACATTCAGCTCTCTAGTACCA 1619  
Db |||||  
55544 ATCAAAGGGGAAACACAGCGGTTTGAATAGGGGAAACACACATTCAGCTCTCTAGTACCA 55603  
Qy |||||  
1620 ATATGCTGTGAAGACACAGCAATGATGTCAGACATACATGATCTAGCGCAGAGATCA 1679  
Db |||||  
55604 ATATGCTGTGAAGACACAGCAATGATGTCAGACATACATGATCTAGCGCAGAGATCA 55663  
Qy |||||  
1680 ATGCCAGATGGAACCCGAAACCTCTGTTCTGAGTCACTCGGCAA----- 1723  
Db |||||  
55664 ATGCCAGATGGAACCCGAAACCTCTGTTCTGAGTCACTCGGCAAAGGTACGAATTTTGTG 55723  
Qy |||||  
1724 ----- 1723  
Db |||||  
55724 AATCNTGAGGAATTTTGTCTTTTAAATGATGATGATGATGATGATGATGATGATGATGATG 55783  
Qy |||||  
1724 ----- 1723  
Db |||||  
55784 TAATATGCTGCATACAAATGTTAAGAAATATGCATACATGTTTATTATATGCTTTCC 55843  
Qy |||||  
1724 ----- 1723  
Db |||||  
55844 ACTGTTCTCTTACTTATCTTTTGTGATCTCTTTTGTGTGCGTGATGATGATGATGATG 55903  
Qy |||||  
1724 ----- 1723  
Db |||||  
55904 GT 55963  
Qy |||||  
1724 ----- AGTTTT 1729  
Db |||||  
55964 AGACTCATATTATAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 56023  
Qy |||||  
1730 CTCAGCTGAGCATGATATCCAAATATGCTGACCTTCATGACGAGTCTACCCAAAGA 1789  
Db |||||  
56024 CTCAGCTGAGCATGATATCCAAATATGCTGACCTTCATGACGAGTCTACCCAAAGA 56083  
Qy |||||  
1790 AGAAAAAGAGCAAAACTTGAAGTGTCTGTGAAAAACAGACCATGATGATGATGATGATGATG 1849  
Db |||||  
56084 AGAAAAAGAGCAAAACTTGAAGTGTCTGTGAAAAACAGACCATGATGATGATGATGATGATG 56143  
Qy |||||  
1850 CCATGATATGTTGAAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1909  
Db |||||  
56144 CCATGATATGTTGAAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 56203  
Qy |||||  
1910 CTGATGTTCTGACATCAACCGTATTCATCCAAAGCAACTGCTGATGATGATGATGATGATG 1969  
Db |||||  
56204 CTGATGTTCTGACATCAACCGTATTCATCCAAAGCAACTGCTGATGATGATGATGATGATG 56263  
Qy |||||  
1970 TAGTAGCTGCCAAGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2029  
Db |||||  
56264 TAGTAGCTGCCAAGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 56323  
Qy |||||  
2030 AGAAGTCTCTGGCATCCCAAGTACACAGAGAGAGTACAGGGTCAATTTGGCAATTTGACCA 2089  
Db |||||  
56324 AGAAGTCTCTGGCATCCCAAGTACACAGAGAGAGTACAGGGTCAATTTGGCAATTTGACCA 56383  
Qy |||||  
2090 CACAGAGTCTCCCATCTCTCAGAACTTTTCACTCTAGTCTAGTCTAGTCTAGTCTAGTCTAG 2149  
Db |||||  
56384 CACAGAGTCTCCCATCTCTCAGAACTTTTCACTCTAGTCTAGTCTAGTCTAGTCTAGTCTAG 56443  
Qy |||||  
2150 GGATGGAAGAAATGGTCTACTATTCGCAAGTCTCAGCACTATTTTCACTCATGATGATG 2209  
Db |||||  
56444 GGATGGAAGAAATGGTCTACTATTCGCAAGTCTCAGCACTATTTTCACTCATGATGATG 56503  
Qy |||||  
2210 AGTATATTGCTGAAGACCAACTGAACTGAACTGCGGCGCGTGAAGACGCAAAAGAGCTAACGT 2269



VERSION	GI:2724471
KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 7218)
AUTHORS	Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE	Recombinant fowlpox virus
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;
FEATURES	Location/Qualifiers
source	1..7218
	/organism="unknown"
	/mol_type="unassigned DNA"
ORIGIN	
Query Match	2.0%; Score 62.8; DB 6; Length 7218;
Best Local Similarity	7.2%; Pred. No. 1.9e-05;
Matches	31; Conservative 227; Mismatches 174; Indels 0; Gaps 0;
Qy	942 ATGTGACGAGATCTGCAATGAGCCATGTGCAAGAAGTTGTTCTCAAAAGAAGTCTCCAAATC 1001
Db	1497 ATGTAGCGATCACTGTATTACCTATCTATGCAAGTAGTTAAAGAGATAGAGAATTGG 1438
Qy	1002 TAAGAGGAAGACGGATAAGAAGTTGATGAAGACGACGACGACGACGACGACGACGACGACGAC 1061
Db	1437 TACER 1378
Qy	1062 CCAGGCTGATGTTTCAGATGCAAGCTTTGTGCGAAGACGACGACGACGACGACGACGACGAC 1121
Db	1377 RER 1318
Qy	1122 ATCAGAAATTATAATGCTTAAACAGTTGAGGATTTCTAGAAGTCACGAAGTTCATCGTGA 1181
Db	1317 RER 1358
Qy	1182 AAATGCGCTGATCCCTGTGAGATGATAGATACCATCCCGTCCCGATGGAAGTAAG 1241
Db	1257 RER 1198
Qy	1242 CATGGATATTCCTGTAGCAACCATCAGTGGGAGAGATGGGTTAAATCAAGTAAGAA 1301
Db	1197 RER 1138
Qy	1302 CAAGACAAAACGCAATACTCTGATGTTGATGATGATGATGATGATGATGATGATGATGATG 1361
Db	1137 RER 1078
Qy	1362 GAATGGAAAAA 1373
Db	1077 RER
RESULT 7	
AX655393/c	2000 bp DNA linear PAT 22-MAR-2003
LOCUS	AX655393
DEFINITION	Sequence 5263 from Patent WO03008096.
ACCESSION	AX655393
VERSION	AX655393.1 GI:29158207
KEYWORDS	
SOURCE	Oryza sativa
ORGANISM	Oryza sativa
REFERENCE	1
AUTHORS	Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M., Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
TITLE	Plant genes involved in defense against pathogens
JOURNAL	Patent: WO 03008096-A 5263 03-JAN-2003;
FEATURES	Syngeta Participations AG (CH)
source	Location/Qualifiers
	1..2000
	/organism="Oryza sativa"

REFERENCE 1  
AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M., Katagiri, F., Qian, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.  
TITLE Plant genes involved in defense against pathogens  
JOURNAL Patent: WO 0300898-A 5263 03-JAN-2003;  
Syngenta Participations AG (CH)  
FEATURES Location/Qualifiers  
source  
1..2000  
/organism="Oryza sativa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:4530"

ORIGIN

Query Match 1.6%; Score 52; DB 6; Length 2000;  
Best Local Similarity 10.0%; Pred. No. 0.015;  
Matches 60; Conservative 273; Mismatches 268; Indels 1; Gaps 1;

QY 280 AGTTCAAGCCCAATTTTCTGTAGCAAGTTTTCAGCAGTGGGATTTGCTCGAAGTCTTGGAT 339  
DB 316 RGYSRMAMMYKQWYRGYKGMGRWAGRMWMSRMSCEWSKACYMRWRMWRMWRMWRMWR 375  
QY 340 AGTTCAAGCCCAATTTTCTGTAGCAAGTTTTCAGCAGTGGGATTTGCTCGAAGTCTTGGAT 399  
DB 376 AKSSRTSRKKRKKMKRKYKRMGRYSRMSRCKGRWKRCKRSGRAWKRCGRGOWTCR 435  
QY 400 ACAAGTGTAGTTTCTGCTCCATCATTTTGTTCGAGCAGTCTTGTGCTGCTGCTGCTGCTG 459  
DB 436 MKSYGMWRKSKWASAKYKMSBMYRWRKCCSRRTYMGKTEGGMWGTGRCYKCRSG 495  
QY 460 TCCAAAAGTGTCTCTAGCAGCAATCATCTCAAGGAGGAAGTCTGCTGATGATCACT 519  
DB 496 MKRCKRRRWRMGRMYRMRKRYTMSRYTMYRCARKEYSYSAARKARCWYRGKGYWAGMW 555  
QY 520 CTTCACAGAGTGTGCAAGAGCAATGACTCCAAATGCAATGCGCTTCTGCGCAAGAT 579  
DB 556 MKRYKRYMYKQWYRKYKSKWYKSKSYKSYKSYKSYKSKYKSKYKSKYKSKYKSKYKSK 615  
QY 580 GGAGCTGTGAGCCCAATTTGATTCACCAATGAAAGATTGCAAGGCGGCGAGCCCAAT 639  
DB 616 SSKCKRCKASXSSAKRYAMMGMTSGSRMSKSYTCYWRKWSKSKSTCTWYMYMSKYT 675  
QY 640 TATGATGTGCGCAATTTCTCTGAGGACACACATCTCTGCTGCTGCTGCTGCTGCTGCT 699  
DB 676 YAKYGYRY 735  
QY 700 GAAGTTTCCCAAGTATCATGCGCAGCATAGAGTAAATGGTGAGATCAACTCACTCACT 759  
DB 736 YCKSKSKYRSMYKSYKSYKSYKSYKSYKSYKSYKSYKSYKSYKSYKSYKSYKSYKSY 795  
QY 760 CCAAACTTTCTG-AGTGTCTCTCAAGAGAAATGAAGATGAAGATGAAGATGAAGATGAAG 818  
DB 796 YMDRYMYKCTKTYWYSATYWTGTAAAWMAKTKTJRMGTGAKTRGRARKARYMKA 855  
QY 819 GACTCTTGTGTCGAGCAGTCAATTTGACCAAGATCTTAACCAATGCTGCTGCAAGGA 878  
DB 856 TWCAKRWMTGKGAARWMTMAKARKYKSYKSYKSYKSYKSYKSYKSYKSYKSYKSYK 915  
QY 879 AC 880  
DB 916 AY 917

RESULT 9  
EX897685/c  
LOCUS Danio rerio clone CH211-234H8, WORKING DRAFT SEQUENCE, 11 unordered  
DEFINITION pieces.  
ACCESSION BX897685  
VERSION BX897685.1 GI:40067210  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

ORIGIN

Query Match 2.0%; Score 62; DB 6; Length 2000;  
Best Local Similarity 10.5%; Pred. No. 2.8e-05;  
Matches 82; Conservative 360; Mismatches 325; Indels 11; Gaps 3;

QY 1664 TCAGCGCAGAGATCAATGCGAGTGAAGAAACCGAAACTCTGTCTGAGTCACTGCGCAA 1723  
DB 939 WCYCCMKWKWMTSCMKWYRTWTCWYTWGAMRYAYVYMRWRWYKWSWRWYK 880  
QY 1724 AGTTTCTCAGCTGAGCAGTATCAATTAATTTGCTGACCTTCATGAGCAGAGCTAC 1783  
DB 879 MTQWATWMTCMKAWYATGATWMTWYTYCYATCACTKCYKAMTAKWTTWAC 820  
QY 1784 CCAAGAAAGAAAGCAAACTTCAAGTGTCTGAGTGAAGAAACCGACCATGATGATAG 1843  
DB 819 AWRATSWRWAAG-NEWKRYKRYRYRYRYRYRYRYRYRYRYRYRYRYRYRYRYRYRY 764  
QY 1844 ACATCCCATGATATTTGTTGAACTGCTAGTAAACACGAGCAGCAGCAGTATGA 1903  
DB 763 YVWQWAMTWMSRRWKSYSRMSGMRMSAWRYSRMKCAKTKYASSARWTFKRS 704  
QY 1904 CTGAGCTGATTTCTGACATCAACCGTATTCAATCAAGACAACTGCTGATGATGAT 1963  
DB 703 YRYRRWYKWKHTYRY 644  
QY 1964 GTGTAATAGTGTGCAAGATGGTTCAGTATGATGATGATGATGATGATGATGATGAT 2023  
DB 643 SCSAKCKCTRYMTSSYMTGMYSYKSMSTSKSYMGKTKTCTMYTSMKGRSRK 584  
QY 2024 C-----CAAAGAGTCTTGGCATCCCAAGTACACAGAGAGTTCAGAGGTCAATT 2078  
DB 583 MGRWSGRMTRWKKRKKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 524  
QY 2079 GGCAATTGACCAACAGAGTCTCCATCTCTCAGAACTTTTCACTGCTACTCAGGAACGA 2138  
DB 523 ARYTSKRYMYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 466  
QY 2139 GACATTTGCGGATGGAAGAAATGCTCATTTGCTGCAAGTCCACCATTATTTTCA 2198  
DB 465 MYRYKYSKRYMYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 406  
QY 2199 TCATGATGATCAGTATATTTGCTGAGCAGCACTCAAGTATGCGGCGCTGAAGCAGCAA 2258  
DB 405 YSRCKYKRYMYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 346  
QY 2259 GAAGCTAAGTGGAGCAATTTAAGGCGCACTCAAGAAATTTCCAGCAGCAACATGCG 2318  
DB 345 CTWCMYKRCYRWRKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 286  
QY 2319 TGCTCAATTTAGACTGTATCCAGCAGTGTGCTGACTTCTACTCATGCTGCTGCTGCT 2378  
DB 285 WKMSRWRKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 226  
QY 2379 TTCCAGCAATTTATGATCTCCCAACCAAGTAAATTTGCGCCACTGAGCAGCAGTATG 2436  
DB 225 MRACRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRYKRY 168

RESULT 8  
AX655393  
LOCUS Sequence 5263 from Patent WO0300898.  
DEFINITION AX655393.1 GI:29158207  
ACCESSION AX655393  
VERSION AX655393.1  
KEYWORDS Oryza sativa  
SOURCE Oryza sativa  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

[illegible]



TITLE	JOURNAL
REFERENCE	AUTHORS
TITLE	JOURNAL

REFERENCE	AUTHORS
TITLE	JOURNAL

COMMENT
---------

TITLE JOURNAL REFERENCE	AUTHORS TITLE JOURNAL
	REFERENCE AUTHORS TITLE JOURNAL
	COMMENT



D	b	26159	ATGATAATCAAAATAAATGAATCATATAATAATGATATTGATGATGACGATTAATTAATAACA	26218
Q	y	1427	GCACAAAAAGTGACACCACCTGGCAGTACTCACGATGATGATGAGAATGATCTGAAA	1486
D	b	26219	ATAACAATAACAATAACATCAACAACATTAAATAATAATGATGATGATGATGATGATGATA	26278
Q	y	1487	ATGGCTTTGACACAATAATGCGATAGACAGAGTCGTGTCAGCATGTATCAGNAATCTCCA	1546
D	b	26279	ATAATAATAATCATATAATAATATGATCATATGATGATGATAAACCAATGATTAATACCAATCA	26338
Q	y	1547	CACAGAGGTGCTCATCAAAGGGGAAAAACACGCCGGTTTCAGTAGGGGAAAAACACATTCCAG	1606
D	b	26339	ATAATAATAATAATAATGATGATTAATCAATAATGATGATGATGATGATGATTAATAATAACA	26398
Q	y	1607	CTGCTAGTAC	1616
D	b	26399	CAACTACAAC	26408

RESULT 14  
LOCUS AE014820/c

DEFINITION Plasmodium falciparum 3D7 chromosome 14 section 5 of 13 of the complete sequence.

ACCESSION AE014820

SOURCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

KEYWORDS AE014820.1 GI:23497305

ORGANISM Plasmodium falciparum 3D7

REFERENCE 1 (bases 1 to 250029)  
AUTHORS Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W., Carlson,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T., James,K., Eisen,J.A., Nene,V., Shallow,S.J., Suh,B., Peterson,J., Kyes,S., Chan,M.-S., Hall,N., Allen,J., Selengut,J., Haft,D., Angiuoli,S., Pertea,M., Martin,D.M.A., Fairlamb,A.H., Mather,M.W., Vaidya,A.B., Ralph,S.A., McFadden,G.I., Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I., Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C., Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and Barrall,B.

TITLE Genome sequence of the human malaria parasite Plasmodium falciparum

JOURNAL Nature 419 (6906), 498-511 (2002)

PUBMED 12368864

REFERENCE 2 (bases 1 to 250029)  
AUTHORS Gardner,M.J.  
TITLE Direct Submission  
JOURNAL Submitted (13-SEP-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES source Location/Qualifiers

- 1..250029 /organism="Plasmodium falciparum 3D7"
- /mol\_type="genomic DNA"
- /isolate="3D7"
- /db\_xref="taxon:36329"
- /chromosome="14"
- 83..122 /rpt\_type=tandem
- 307..346 /rpt\_type=tandem
- 345..442 /rpt\_type=tandem
- 575..646 /rpt\_type=tandem
- <833..>3733 /locus\_tag="PF14\_0238"
- <833..>3733 /locus\_tag="PF14\_0238"
- 833..3733 /locus\_tag="PF14\_0238"
- /codon\_start=1
- /product="hypothetical protein"
- /protein\_id="FAAN36851.1"

CDS





## COMMENT

On Apr 5, 2000 this sequence version replaced gi:7264205.  
All repeats were identified using RepeatMasker;  
Smit, A.P.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7353

Center clone name: 78 O 22

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 132847 bases at least Q40

Consensus quality: 137397 bases at least Q30

Consensus quality: 139005 bases at least Q20

Insert size: 145000; agarose-fp

Insert size: 139941; sum-of-contigs

Quality coverage: 4.8 in Q20 bases; agarose-fp

Quality coverage: 4.9 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

```

* 1
* 2190: contig of 2190 bp in length
* 2191: 2290: gap of 100 bp
* 2291: 5642: contig of 3352 bp in length
* 5643: 5742: gap of 100 bp
* 5743: 10058: contig of 4316 bp in length
* 10059: 10158: gap of 100 bp
* 10159: 14819: contig of 4661 bp in length
* 14820: 14919: gap of 100 bp
* 14920: 19408: contig of 4489 bp in length
* 19409: 19509: gap of 100 bp
* 19509: 26115: contig of 6607 bp in length
* 26116: 26215: gap of 100 bp
* 26216: 39655: contig of 13440 bp in length
* 39656: 39755: gap of 100 bp
* 39756: 54129: contig of 14374 bp in length
* 54130: 54229: gap of 100 bp
* 54230: 64518: contig of 10289 bp in length
* 64519: 64618: gap of 100 bp
* 64619: 79904: contig of 15286 bp in length
* 79905: 80004: gap of 100 bp
* 80005: 109845: contig of 29841 bp in length
* 109846: 109945: gap of 100 bp
* 109946: 141041: contig of 31096 bp in length.

```

## FEATURES

```

source
1..141041
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-78022"
/clone_lib="RPC1-11 Human Male BAC"
1..2190
/note="assembly_fragment"
2291..5642
/note="assembly_fragment"
5743..10058
/note="assembly_fragment"
10159..14819
/note="assembly_fragment"
14920..19408

```

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

```

/note="assembly_fragment"
19509..26115
/note="assembly_fragment"
26216..39655
/note="assembly_fragment"
39756..54129
/note="assembly_fragment"
54230..64518
/note="assembly_fragment"
64619..79904
/note="assembly_fragment"
80005..109845
/note="assembly_fragment"
109946..141041
/note="assembly_fragment"

```

## ORIGIN

```

Query Match      1.5%; Score 48.2; DB 2; Length 141041;
Best Local Similarity 46.5%; Pred. No. 0.25;
Matches 155; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

QY 1178 GTGAAATGCGGTGATCCCTGTGAGGATGATAGAAATACCATCCCGGTCCCGATCGAAG 1237
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98342 GAGATCATGCGCATTTACACTCCAGCTGGTCAACAGAGCGAGATTCCATCTCAAAAAAAA 98401
QY 1238 TAACATGGATATTCCTGTTAGCAACCATACATACATGGAGAGATGGTTAAATCAAGTA 1297
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98402 AAAAAAGTAAATGGCATTAGGAAAAACATTGTAAAGGAGAGAGGTAATATAGTTACAA 98461
QY 1298 AGAACAAAGACAAACGCAAAATCTCTGTGATGTTGTAGATGATGATCATCATCTTATGAAC 1357
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98462 ATAGAAATCAAAACCAACCAACAGTTAGACTTAAAAAAGATATTTAAAAAGATTGAAAC 98521
QY 1358 GGCTGAATGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1417
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98522 CTGATGGAGAAATCTGTAAAGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 98581
QY 1418 GGAATTTGAGCAACAAAAAGTACACCCCACTCCGAGTACTCAGCATGATGATGAGATG 1477
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98582 GACATTGAAGCAACAAAAAGTACACAAAAAATTTAAAAAATTACAAATTATCAAAATATTCT 98641
QY 1478 ATACTGAAATGGTCTTGACACAAATATGCATA 1510
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98642 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 98674

```

Search completed: May 16, 2004, 23:34:15  
Job time : 12119 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: May 16, 2004, 15:00:26 ; Search time 1184 Seconds  
(without alignments)  
11377.564 Million cell updates/sec

Title: US-09-721-114-1\_COPY\_655\_3825  
Perfect score: 3171  
Sequence: 1 agggacatgttcagtaga.....acgagtatatggattaccgc 3171

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001as:\*  
5: Geneseq2001bs:\*  
6: Geneseq2002s:\*  
7: Geneseq2003as:\*  
8: Geneseq2003bs:\*  
9: Geneseq2003cs:\*  
10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3171	100.0	3896	7 ABQ77908	Abq77908 Rice OsEM
2	3171	100.0	4310	6 ABA93084	Abq93084 Oryza sat
3	2573.2	81.1	9455	6 ABA93085	Abq93085 Oryza sat
4	73.4	2.3	279	6 ABL73385	Ab173385 Corn tass
5	62	2.0	2000	7 ADA71938	Ada71938 Rice gene
6	52	1.6	2000	7 ADA71938	Ada71938 Rice gene
7	45.4	1.4	274	7 ABA85499	Abx85499 Corn ear-
8	42.8	1.3	574	3 AAC41195	Aac41195 Arabidops
9	42.2	1.3	110000	6 ABQ69245_18	Continuation (19 o
10	42.2	1.3	319630	6 ABQ67194	Abq67194 Listeria
11	41.4	1.3	246	5 AAS90679	Aas90679 DNA encod
12	41.4	1.3	1269	6 ABN90734	Abn90734 Staphyloc
13	41.4	1.3	2456	5 AAS75925	Aas75925 DNA encod
14	41.4	1.3	2848	4 AAH54196	Aah54196 S. epider
15	41	1.3	1260	7 ACA46931	Ac446931 Prokaryot
16	40.6	1.3	297	6 ABN90752	Abn90752 Staphyloc
17	40.6	1.3	5299	7 AAD49594	Aad49594 Human cyt
18	40.6	1.3	5322	7 AAD49594	Aad49594 Human cyt
19	40.6	1.3	5451	9 ADC26234	Adc26234 Human NOV
20	40.6	1.3	5761	6 AB211141	Ab211141 Human pol
21	40.4	1.3	686	3 AAO9318	Aao9318 Human can
22	40.4	1.3	110000	6 ABA03041_17	Continuation (18 o
23	40	1.3	412	8 ACH35151	Ach35151 Human end

C	24	39.8	1.3	15122	6 ABL32775	Ab132775 Human imm
C	25	39.6	1.2	3257	4 AAH54703	Aah54703 S. epider
C	26	39.4	1.2	843	5 AAS92080	Aas92080 DNA encod
C	27	39.2	1.2	2522	2 AAV63196	Aav63196 cDNA from
C	28	39.2	1.2	2522	6 ABQ92046	Abq92046 Human pol
C	29	39.2	1.2	6065	6 ABL32504	Ab132504 Human imm
C	30	38.8	1.2	648	7 ABT42404	Abt42404 Toxicity
C	31	38.8	1.2	648	9 ADB59062	Adb59062 Toxicity
C	32	38.8	1.2	648	9 ADB53815	Abd53815 Primary r
C	33	38.8	1.2	3291	7 ABZ42128	Abz42128 Arabidops
C	34	38.8	1.2	14041	4 AAH48024	Aah48024 Internal
C	35	38.6	1.2	261	4 AAI27301	Aai27301 Probe #17
C	36	38.6	1.2	261	4 AAI22038	Aai22038 Probe #11
C	37	38.6	1.2	261	4 ABA67114	Ab67114 Human foe
C	38	38.6	1.2	261	4 ABA75585	Ab75585 Human foe
C	39	38.6	1.2	261	4 AAI56167	Aai56167 Probe #24
C	40	38.6	1.2	261	4 AAI47330	Aai47330 Probe #16
C	41	38.6	1.2	261	4 ABA49199	Ab49199 Human bre
C	42	38.6	1.2	261	4 ABA34204	Ab34204 Probe #12
C	43	38.6	1.2	261	4 ABA40197	Ab40197 Probe #18
C	44	38.6	1.2	261	4 AAK41287	Aak41287 Human bon
C	45	38.6	1.2	261	4 AAK50216	Aak50216 Human bon

ALIGNMENTS

RESULT 1  
ABQ77908  
ID ABQ77908 standard; cDNA; 3896 BP.  
AC ABQ77908;  
XX  
XX  
DT 24-JAN-2003 (first entry)  
DE Rice OsEMF1-encoding cDNA, SEQ ID NO:1.  
XX  
XX  
KW Rice; OsEMF1; EMF; embryonic flower; plant; floral repressor;  
KW reproductive development; flower development; transgenic plant;  
KW antisense suppression; transgenic; reciprocal negative interaction;  
KW flower meristem identity gene; flowering time; shoot development;  
KW seed yield; agriculture; gene; ss.  
OS Oryza sativa.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 241..3414  
FT FT /\*tag= a  
FT FT /product= "OsEMF1 protein"

XX WO200280659-A1.  
XX  
XX 17-OCT-2002.  
XX  
XX 05-APR-2002; 2002WO-US012675.  
XX  
XX 06-APR-2001; 2001US-00828068.  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Moon Y, Chen L, Sung ZR;  
XX WPI; 2003-046831/04.  
XX P-PSDB; ABB99878.  
XX  
XX New isolated OsEMF1 nucleic acid molecule and encoded polypeptide, useful  
XX for plant genetic engineering, in particular controlling reproductive  
XX development in rice.  
XX  
XX Claim 3; Fig 1; 47pp; English.  
XX  
XX The invention relates to a rice EMF (embryonic flower) gene designated  
XX OsEMF1 (cDNA given in ABQ77908) and its encoded protein (ABB99878). The



Db 1981 CATGATATCCAAATTAATGCTGACCTTCAATGACGAGTCTACCCAGAGAGAAAAGAG 2040  
Qy 1801 CAAAACTTGAAGTCACTGCTGAAAGAAACAGACCATGATAGATGACATCCCATGATAT 1860  
Db 2041 CAAAACTTGAAGTCACTGCTGAAAGAAACAGACCATGATAGATGACATCCCATGATAT 2100  
Qy 1861 GTTGAATGCTAGCTAATAAAACAGCAATGAGAGGAGCTTATGACTGAGCTGATGTTCT 1920  
Db 2101 GTTGAATGCTAGCTAATAAAACAGCAATGAGAGGAGCTTATGACTGAGCTGATGTTCT 2160  
Qy 1921 GACATCAACCGTATTAATCAATCCCAAGACAACTGCTGATGATGATGTTGTTAATAGTAGCTGCC 1980  
Db 2161 GACATCAACCGTATTAATCAATCCCAAGACAACTGCTGATGATGATGTTGTTAATAGTAGCTGCC 2220  
Qy 1981 AAGGATGCTAGCTAATAAAACAGCAATGAGAGGAGCTTATGACTGAGCTGATGTTCT 2040  
Db 2221 AAGGATGCTAGCTAATAAAACAGCAATGAGAGGAGCTTATGACTGAGCTGATGTTCT 2280  
Qy 2041 GCATCCCAAGTACACAGAGAGGATGACAGGCTCAATTTGGCAATGACCAACAGAGTCT 2100  
Db 2281 GCATCCCAAGTACACAGAGAGGATGACAGGCTCAATTTGGCAATGACCAACAGAGTCT 2340  
Qy 2101 CCACATCTCTCAGAACTTTTCACTCTACTCAGAGAACAGCAGACACATTTGGGATGGAAGAA 2160  
Db 2341 CCACATCTCTCAGAACTTTTCACTCTACTCAGAGAACAGCAGACACATTTGGGATGGAAGAA 2400  
Qy 2161 ATGGTCACTATTTGCGAAGCTCAACATATTTTCACTATCATGATGATGATGATGATGCT 2220  
Db 2401 ATGGTCACTATTTGCGAAGCTCAACATATTTTCACTATCATGATGATGATGATGCT 2460  
Qy 2221 GAAGCACAACCTGAACTTTGGGCGCTTAAGGACGCAAGAGCAATTAACGTTGGAGCAATTT 2280  
Db 2461 GAAGCACAACCTGAACTTTGGGCGCTTAAGGACGCAAGAGCAATTAACGTTGGAGCAATTT 2520  
Qy 2281 AAGGCCACTACAGAAATTTCTCCAGCAGCAAACTGTTGCTCAATTTAGACCTGGTATC 2340  
Db 2521 AAGGCCACTACAGAAATTTCTCCAGCAGCAAACTGTTGCTCAATTTAGACCTGGTATC 2580  
Qy 2341 CAAGCAGTTGACTGCTGCTTCTACTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 2400  
Db 2581 CAAGCAGTTGACTGCTGCTTCTACTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 2640  
Qy 2401 CAACCAAGTAAATTTGGGCCACTGCGCACTGCGCACTGCGCACTGCGCACTGCGCACTGCG 2460  
Db 2641 CAACCAAGTAAATTTGGGCCACTGCGCACTGCGCACTGCGCACTGCGCACTGCGCACTGCG 2700  
Qy 2461 AGAAATTTCCAGGCAATAGCAACCATGAGGAGTAACTGATGATGATGATGATGATGATGAT 2520  
Db 2701 AGAAATTTCCAGGCAATAGCAACCATGAGGAGTAACTGATGATGATGATGATGATGATGAT 2760  
Qy 2521 GCTGACCAAGTATGCTTGTATCTTAAGAAATCCATGCTGCGACGCACTTCTCTGGAATG 2580  
Db 2761 GCTGACCAAGTATGCTTGTATCTTAAGAAATCCATGCTGCGACGCACTTCTCTGGAATG 2820  
Qy 2581 ATGGATCCATCAACATAGCAAGCTTCCCACTATGGAATCTTAGCAGGAAACAGATG 2640  
Db 2821 ATGGATCCATCAACATAGCAAGCTTCCCACTATGGAATCTTAGCAGGAAACAGATG 2880  
Qy 2641 GAGTCTCAATTCATTAATTTCTCAGTATGACATTAATCAGTACAAAGGATCAACAGGACCA 2700  
Db 2881 GAGTCTCAATTCATTAATTTCTCAGTATGACATTAATCAGTACAAAGGATCAACAGGACCA 2940  
Qy 2701 TCATATGGCAGTAACCTGATGAAAGATTTCCATGACATTCGAGCACTTATCAAGCAATGAT 2760  
Db 2941 TCATATGGCAGTAACCTGATGAAAGATTTCCATGACATTCGAGCACTTATCAAGCAATGAT 3000  
Qy 2761 CAGTCTCATGATCTGACAGACCTTTACGCCCATCTTACGCCCATCTTACGCCCATCTTACGCC 2820  
Db 3001 CAGTCTCATGATCTGACAGACCTTTTACGCCCATCTTACGCCCATCTTACGCCCATCTTACGCC 3060  
Qy 2821 TTGCTGCAAGGAAATTTGCAATCTGCTCGGAACTGTCGACCAACAAATCTGTTTATAG 2880  
Db 3061 TTGCTGCAAGGAAATTTGCAATCTGCTCGGAACTGTCGACCAACAAATCTGTTTATAG 3120

## RESULT 2

ABA93084

ID ABA93084 standard; cDNA; 4310 BP.

XX ABA93084;

AC ABA93084;

XX 11-APR-2002 (first entry)

DT 11-APR-2002 (first entry)

XX Oryza sativa brassinosteroid response related protein encoding cDNA.

DE Oryza sativa; rice; plant; brassinosteroid response; control;

XX signal transfer system; brassinosteroid hormone; growth promotion;

KW increased yield; quality improvement; ripeness promotion;

KW stress relaxation; chemical resistance; gene; ss.

XX Oryza sativa.

XX Key

FH 655..3828

FT CDS

FT Location/Qualifiers

XX /\*tag= a

FT /product= "brassinosteroid response related protein"

XX JP2001327287-A.

XX 27-NOV-2001.

XX 19-MAY-2000; 2000JP-00149106.

XX 19-MAY-2000; 2000JP-00149106.

XX (NORQ) NORINSUISANSHO NOGYO SEIBUTSU SHIGEN.

XX (SEIB-) SEIBUTSUKI TOKUTEI SANGYO GIJUTSU.

XX WPI; 2002-135739/18.

XX P-PSDB; ABB05428.

XX New gene involved in brassinosteroid responses useful for controlling the

XX effects such as growth promotion, increased yield, quality improvement,

XX ripeness promotion, stress relaxation and chemical resistance.

XX Disclosure; Page 6-10; 19pp; Japanese.

XX The present invention describes a polynucleotide encoding a plant gene

XX which can control the signal transfer system of brassinosteroid hormone.

XX The polynucleotide can be used for controlling the effects such as growth

XX promotion, increased yield, quality improvement, ripeness promotion,

XX stress relaxation and chemical resistance. The present sequence encodes a

XX brassinosteroid response related protein isolated from rice (Oryza

XX sativa), from the present invention

XX Sequence 4310 BP; 1237 A; 986 C; 1022 G; 1065 T; 0 U; 0 Other;

SQ

Qy 2881 TTAGAGTGTCAACAGGAATAACATCGCATCAGATGAACAGAAAGGAACATTTTGAAGCC 2940  
Db 3121 TTAGAGTGTCAACAGGAATAACATCGCATCAGATGAACAGAAAGGAACATTTTGAAGCC 3180  
Qy 2941 CTGAATTTCTGAAATGTTTTCAGCAAAATGGAATGCAATTCGATTTGGTTCTGTTAGCTCC 3000  
Db 3181 CTGAATTTCTGAAATGTTTTCAGCAAAATGGAATGCAATTCGATTTGGTTCTGTTAGCTCC 3240  
Qy 3001 AGTCGAGATTTTATCAGCGAGGAACAGCATAGCTCAATCTTGAACAGAGGCAAGGGT 3060  
Db 3241 AGTCGAGATTTTATCAGCGAGGAACAGCATAGCTCAATCTTGAACAGAGGCAAGGGT 3300  
Qy 3061 AAAATGTTTATCCTCGGATCGGTTTGTGAGCAGGATATCTGTATTAACATAACAGAAC 3120  
Db 3301 AAAATGTTTATCCTCGGATCGGTTTGTGAGCAGGATATCTGTATTAACATAACAGAAC 3360  
Qy 3121 CCAGTCGATTTTACTACAATCAGTAAACGATAACGATATATGATATACCGC 3171  
Db 3361 CCAGTCGATTTTACTACAATCAGTAAACGATAACGATATATGATATACCGC 3411

Query Match	100.0%;	Score 3171;	DB 6;	Length 4310;
Best local Similarity	100.0%;	Pred. No. 0;		
Matches 3171;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATCGAGATTGTTTCAGTAGATCAGGAGGGAGCTCGTGTGTTGTTGGGACGAACTGTATGCTT	60	
DB	655	ATCGAGATTGTTTCAGTAGATCAGGAGGGAGCTCGTGTGTTGTTGGGACGAACTGTATGCTT	714	
QY	61	GCTCGTGGTGGAACTGGTGTCTGTAGGCGCAGTGTTCGAGCTGCACGAGCGCTCCTGTCAG	120	
DB	715	GCTCGTGGTGGAACTGGTGTCTGTAGGCGCAGTGTTCGAGCTGCACGAGCGCTCCTGTCAG	774	
QY	121	GATCGAGCGCTGAAGCTGTGTAGAGGAACCGGCAACAACCAATGTGCGAGCAATTTCTCC	180	
DB	775	GATCGAGCGCTGAAGCTGTGTAGAGGAACCGGCAACAACCAATGTGCGAGCAATTTCTCC	834	
QY	181	ATAAGAGGGTATGTTGCTCTTCTTCAGAGGAAGATCCAAAATTTCTGCTCTATCTCGG	240	
DB	835	ATAAGAGGGTATGTTGCTCTTCTTCAGAGGAAGATCCAAAATTTCTGCTCTATCTCGG	894	
QY	241	ATTTTCATCACCACAGAAAAATGTGATGAACAACAAGCTAGTTCAAGCCCATTTTCTGTA	300	
DB	895	ATTTTCATCACCACAGAAAAATGTGATGAACAACAAGCTAGTTCAAGCCCATTTTCTGTA	954	
QY	301	GCAAAAGTTTCGACGATGGGATGCTCGAAGTGTCTGGATAAGTTGAAAACCTTCAGATAAT	360	
DB	955	GCAAAAGTTTCGACGATGGGATGCTCGAAGTGTCTGGATAAGTTGAAAACCTTCAGATAAT	1014	
QY	361	GGACACGACCAAGAACTCTTCGGCAACAGCAAGTGGCAACAGTGTGTTGCTCCATC	420	
DB	1015	GGACACGACCAAGAACTCTTCGGCAACAGCAAGTGGCAACAGTGTGTTGCTCCATC	1074	
QY	421	ACATTTGTTCCGGAGCACTTTTGTCGCTGCTAGTGTGTTGTTTCCAAAAAGTGTCTCCTAGC	480	
DB	1075	ACATTTGTTCCGGAGCACTTTTGTCGCTGCTAGTGTGTTTCCAAAAAGTGTCTCCTAGC	1134	
QY	481	ACAAATCATCTCAAGGGAAGATGCTGATAGATCAACTCTTCCAAAGAGTGTGCGAAGAA	540	
DB	1135	ACAAATCATCTCAAGGGAAGATGCTGATAGATCAACTCTTCCAAAGAGTGTGCGAAGAA	1194	
QY	541	GGCAATGACTCCAAATGCAATGCGCCTTCTCGCAAGATGGAGCTGCTGAGGCCAATACT	600	
DB	1195	GGCAATGACTCCAAATGCAATGCGCCTTCTCGCAAGATGGAGCTGCTGAGGCCAATACT	1254	
QY	601	GATTCCAAATGAAAGATTTGCAAGGCCAGCCCAAAATATGATGTGGCAGCAAAATGTC	660	
DB	1255	GATTCCAAATGAAAGATTTGCAAGGCCAGCCCAAAATATGATGTGGCAGCAAAATGTC	1314	
QY	661	TCTGAGGACAACTCTCTGTTGATGTTGGGCTTTTACCTGAAGTTCCCAAGATTACATGG	720	
DB	1315	TCTGAGGACAACTCTCTGTTGATGTTGGGCTTTTACCTGAAGTTCCCAAGATTACATGG	1374	
QY	721	CACATAGAAGTAAATGTTGCAGATCAACTCCATCCACTCCAAAACCTTCTGAAGTGGTC	780	
DB	1375	CACATAGAAGTAAATGTTGCAGATCAACTCCATCCACTCCAAAACCTTCTGAAGTGGTC	1434	
QY	781	CTCAAAAGAAATGAAATGAAATGAAATGAAATGAAAGACTCTTGTGCTGCGCAGTGC	840	
DB	1435	CTCAAAAGAAATGAAATGAAATGAAATGAAATGAAAGACTCTTGTGCTGCGCAGTGC	1494	
QY	841	AAATTGACCAAGATCTTAACCCCAATGTCTGGAAGGAAGTGCATCAGTTGCTGAGCAG	900	
DB	1495	AAATTGACCAAGATCTTAACCCCAATGTCTGGAAGGAAGTGCATCAGTTGCTGAGCAG	1554	
QY	901	TGCAATTTGACCAAGATCCGAAACCAAGTGTCTGGGCAAGAAATGTGAGCAGTCTCAAT	960	
DB	1555	TGCAATTTGACCAAGATCCGAAACCAAGTGTCTGGGCAAGAAATGTGAGCAGTCTCAAT	1614	
QY	961	GAGCCATGTGAAGAGTTGTTCTCAAAGAGAGCTCCAAATCTAAGAGGAAGACGGAATAAG	1020	
DB	1615	GAGCCATGTGAAGAGTTGTTCTCAAAGAGAGCTCCAAATCTAAGAGGAAGACGGAATAAG	1674	

DB 2755 CCACATCTCTCAGAACTTTTCAGTCTACTCAGGAACAGACACATTTTGGGATGGAAGA 2814  
QY 2161 ATGGTCACTATTGCTCGAAGCTCACCACATTTTTCATCATCATGATCATCATGATATTGCT 2220  
DB 2815 ATGGTCACTATTGCTCGAAGCTCACCACATTTTTCATCATCATGATCATCATGATATTGCT 2874  
QY 2221 GAAGCAACCACTGAACATTTGGGCGCTTAAGAGCGMAAGAGCTAACGTTGGAGCAATTT 2280  
DB 2875 GAAGCAACCACTGAACATTTGGGCGCTTAAGAGCGMAAGAGCTAACGTTGGAGCAATTT 2934  
QY 2281 AAGGCCACTACAGAAATTTCTCCAGCAGCAACATGTTGGTGTCTCAATTTAGACTGCTGATC 2340  
DB 2935 AAGGCCACTACAGAAATTTCTCCAGCAGCAACATGTTGGTGTCTCAATTTAGACTGCTGATC 2994  
QY 2341 CAAGCAAGTTGACTTCTTCTACTCTATGTCATGGGATCTTCCAGCAATTTATGTCATCTGC 2400  
DB 2995 CAAGCAAGTTGACTTCTTCTACTCTATGTCATGGGATCTTCCAGCAATTTATGTCATCTGC 3054  
QY 2401 CAACAGTATATGGCGACTTGGCCGCTATGCTGAAGAGCGGTAAACAGGTCCATGCA 2460  
DB 3055 CAACAGTATATGGCGACTTGGCCGCTATGCTGAAGAGCGGTAAACAGGTCCATGCA 3114  
QY 2461 AGAAATTTTCCAGCAACATAGCAACCATGGAAGCGAGTAAGTTATGTCATCGGAGAAAT 2520  
DB 3115 AGAAATTTTCCAGCAACATAGCAACCATGGAAGCGAGTAAGTTATGTCATCGGAGAAAT 3174  
QY 2521 GCTGGAACAAGTATCTTGAATCTTAAAGAAATCCATGCTGCGACCGCATCTTCTGAGATG 2580  
DB 3175 GCTGGAACAAGTATCTTGAATCTTAAAGAAATCCATGCTGCGACCGCATCTTCTGAGATG 3234  
QY 2581 ATGATTCACATACATAGCAACCTTCCCACTATGGAACCTTCTAGCAGGAACCAAGATG 2640  
DB 3235 ATGATTCACATACATAGCAACCTTCCCACTATGGAACCTTCTAGCAGGAACCAAGATG 3294  
QY 2641 GAGTCTCAACTTCATATTTCTAGTATGCAATATCAATCAAGAAATCAACAGCA 2700  
DB 3295 GAGTCTCAACTTCATATTTCTAGTATGCAATATCAATCAAGAAATCAACAGCA 3354  
QY 2701 TCATATGCAAGTAACTGGAATGGAAGATTCATTCATTCATTCATTCATTCATTCATTCAT 2760  
DB 3355 TCATATGCAAGTAACTGGAATGGAAGATTCATTCATTCATTCATTCATTCATTCATTCAT 3414  
QY 2761 CAGCTGATGATCTGCACAGACCTTTTACGCCACATCTCTAGAGTTGGTGTGCTTGGCTCC 2820  
DB 3415 CAGCTGATGATCTGCACAGACCTTTTACGCCACATCTCTAGAGTTGGTGTGCTTGGCTCC 3474  
QY 2821 TTGCTGCAAGAAATTTGAAACTGGTGGAGAACTGGTGGCAACAATCTGGTTATAG 2880  
DB 3475 TTGCTGCAAGAAATTTGAAACTGGTGGAGAACTGGTGGCAACAATCTGGTTATAG 3534  
QY 2881 TTAGGAGTGTCAACAGGAATAACATCGCATCAATCAAGAAAGAAATTTTGAAGCC 2940  
DB 3535 TTAGGAGTGTCAACAGGAATAACATCGCATCAATCAAGAAAGAAATTTTGAAGCC 3594  
QY 2941 CTGAATTTCTGGAAATGTTTTTCAAGAAATGGAATGCAATTCAGTTGGGTTCTGTTAGCTCC 3000  
DB 3595 CTGAATTTCTGGAAATGTTTTTCAAGAAATGGAATGCAATTCAGTTGGGTTCTGTTAGCTCC 3654  
QY 3001 AGTGCAGATTTTATCAGCAGAGCAACATAGCTCAATCTTGGACCAAGGCAAGGT 3060  
DB 3655 AGTGCAGATTTTATCAGCAGAGCAACATAGCTCAATCTTGGACCAAGGCAAGGT 3714  
QY 3061 AAAATGTTTCTATCCCTTGGATCGGTTTGTGAGCAGGATATCTGTATAACTAACAGAAC 3120  
DB 3715 AAAATGTTTCTATCCCTTGGATCGGTTTGTGAGCAGGATATCTGTATAACTAACAGAAC 3774  
QY 3121 CCAGCTGATTTTACTACATCAGTACGATTAACAGGATATGATGATACGCG 3171  
DB 3775 CCAGCTGATTTTACTACATCAGTACGATTAACAGGATATGATGATACGCG 3825

ABA93085  
ID ABA93085 standard; DNA; 9455 BP.  
XX  
AC ABA93085;  
XX  
DT 11-APR-2002 (first entry)  
XX  
DE Oryza sativa brassinosteroid response related protein related DNA.  
XX  
KW Oryza sativa; rice; plant; brassinosteroid response; control;  
KW signal transfer system; brassinosteroid hormone; growth promotion;  
KW increased yield; quality improvement; ripeness promotion;  
KW stress relaxation; chemical resistance; gene; ds.  
XX  
OS Oryza sativa.  
XX  
PN JP2001327287-A.  
XX  
PD 27-NOV-2001.  
XX  
PF 19-MAY-2000; 2000JP-00149106.  
XX  
PR 19-MAY-2000; 2000JP-00149106.  
XX  
PA (NORQ) NORINSUISANSHO NOGYO SEIBUTSU SHIGEN.  
XX  
PE (SEIB-) SEIBUTSUKI TOKUTEI SANGYO GIJUTSU.  
XX  
WI WIPI; 2002-135739/18.  
XX  
PT New gene involved in brassinosteroid responses useful for controlling the  
PT effects such as growth promotion, increased yield, quality improvement,  
PT ripeness promotion, stress relaxation and chemical resistance.  
XX  
PS Disclosure; Page 13-16; 19pp; Japanese.  
XX  
CC The present invention describes a polynucleotide encoding a plant gene  
CC which can control the signal transfer system of brassinosteroid hormone.  
CC The polynucleotide can be used for controlling the effects such as growth  
CC promotion, increased yield, quality improvement, ripeness promotion,  
CC stress relaxation and chemical resistance. The present sequence  
CC represents a DNA sequence related to a brassinosteroid response related  
CC protein isolated from rice (Oryza sativa), from the present invention  
XX  
SQ Sequence 9455 BP; 2571 A; 2001 C; 2090 G; 2793 T; 0 U; 0 Other;  
Query Match 81.1%; Score 2573.2; DB 6; Length 9455;  
Best Local Similarity 87.8%; Pred. No. 0;  
Matches 3003; Conservative 0; Mismatches 23; Indels 396; Gaps 2;  
QY 146 ACGAACCGGCAACACCAATCGGAGCATTTCTCCATGAAGGGTATGTTGCTTCTTTC 205  
DB 5447 ATGCATCTTTTATAATGATAATCTTATTTTCTGTACAGAGGGTATGTTGCTTCTTTC 5506  
QY 206 AGAACAGGATCCAAATTTCTGCTCTATCTCGATTTTCCATGACCAAGAAATGTTG 265  
DB 5507 AGAAGAGATCCAAATTTCTGCTCTATCTCGATTTTCCATGACCAAGAAATGTTG 5566  
QY 266 ATGAACACAAAGCTAGTTCAAGCCCATTTCTGTAGCAAGTTTTCGAGATGGGATGCT 325  
DB 5567 ATGAACACAAAGCTAGTTCAAGCCCATTTCTGTAGCAAGTTTTCGAGATGGGATGCT 5626  
QY 326 CGAAGTCTTGGATTAAGTTGAATCTTCAGATAATGGAACAGCAACCAAGAACTCTTCCG 385  
DB 5627 CGAAGTCTTGGATTAAGTTGAATCTTCAGATAATGGAACAGCAACCAAGAACTCTTCCG 5686  
QY 386 CAAAGCAGATGGCAACAGTATGTTGCTTCCATCACAATTTTTCGGAGCACTTTTGTGC 445  
DB 5687 CAAAGCAGATGGCAACAGTATGTTGCTTCCATCACAATTTTTCGGAGCACTTTTGTGC 5746  
QY 446 CTGCTAGTGTGGTTCCTCCAAAGTGTCTCTCCAGCACAATCTCTCAAGGAGCAATG 505  
DB 5747 CTGCTAGTGTGGTTCCTCCAAAGTGTCTCTCCAGCACAATCTCTCAAGGAGCAATG 5806



506 CTGATAGATCAACTCTTCCAAAGAGTGTGCAAGAGGCAATGACTCCAAATGCAATGGCG 565  
5807 CTGATAGATCAACTCTTCCAAAGAGTGTGCAAGAGGCAATGACTCCAAATGCAATGGCG 5866  
566 CTTCTGGCAAGAAATGGAGCTGTGAGGCCAAATGACTGATTCACCAATGAA----- 614  
5867 CTTCTGGCAAGAAATGGAGCTGTGAGGCCCAATGACTGATTCACCAATGAAAGGTATGGTAG 5926  
615 ----- 614  
5927 ATGTAGAGCCTTTCAAAATTCCTAAGTAGAGATTTTATTTAAGSTNTAGATAAACTAAATGT 5986  
615 -----AGATTTGCAAGGCCAGCCCAAAATTAATGATGTGGCGAGCAAAATGT 659  
5987 TTGTGTGATTTCTCAGATTTGCAAGGCCAGCCCAAAATTAATGATGTGGCGAGCAAAATGT 6046  
660 CTTCTGAGGACACACCTTTCTGATGTGGGCTTACCTGAAAGTTCCCCAGATTTACATG 719  
6047 CTTCTGAGGACACACCTTTCTGATGTGGGCTTACCTGAAAGTTCCCCAGATTTACATG 6106  
720 GCACATAGAAATGAAATGAGTGCAGATCAACCTCCATCCACTCCAAAACTTTCTGAAAGTGT 779  
6107 GCACATAGAAATGAAATGAGTGCAGATCAACCTCCATCCACTCCAAAACTTTCTGAAAGTGT 6166  
780 CCTCAAAAGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 839  
6167 CCTCAAAAGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 6226  
840 CAATTTGACCAAGATCTTAACCAATGCTGGAAGAAAGTGTGATCAAGTGTGCTGAGCA 899  
6227 CAATTTGACCAAGATCTTAACCAATGCTGGAAGAAAGTGTGATCAAGTGTGCTGAGCA 6286  
900 GTGCAATTTGACCAAGATCTTAACCAATGCTGGAAGAAAGTGTGATCAAGTGTGCTGAGCA 959  
6287 GTGCAATTTGACCAAGATCTTAACCAATGCTGGAAGAAAGTGTGATCAAGTGTGCTGAGCA 6346  
960 TGAGCCATGTAAGAAATGTTCTCAAAAGAGTCTCAAAATCTTAAGAGAAACCGATAA 1019  
6347 TGAGCCATGTAAGAAATGTTCTCAAAAGAGTCTCAAAATCTTAAGAGAAACCGATAA 6406  
1020 GAATTTGATGAAGAACAGACAGACAGCAAGCAAGAAAGCACTGCCAGGCTGATTTTCAGA 1079  
6407 GAATTTGATGAAGAACAGACAGACAGCAAGCAAGAAAGCACTGCCAGGCTGATTTTCAGA 6466  
1080 TGCAAGCTTTGCGAGAAAGCCAAAGAGTGTGGCTTCTATCAGAAATTTATAATGC 1139  
6467 TGCAAGCTTTGCGAGAAAGCCAAAGAGTGTGGCTTCTATCAGAAATTTATAATGC 6526  
1140 TAAACCAAGTTGAGGATTTCTAGAAAGTGCAGAAAGTTCTATCGTGAATAATGCGCGCTGATCCCTG 1199  
6527 TAAACCAAGTTGAGGATTTCTAGAAAGTGCAGAAAGTTCTATCGTGAATAATGCGCGCTGATCCCTG 6586  
1200 TGAGGATGATGAAGTACCATCCCGTCCCGATGGAAGTGAAGTGAATTTCTGTTAG 1259  
6587 TGAGGATGATGAAGTACCATCCCGTCCCGATGGAAGTGAAGTGAATTTCTGTTAG 6646  
1260 CAACCATACAGTGGGAGAGATGGGTTAAATCAAGTAAAGAACAGCAAAACGCAAAATA 1319  
6547 CAACCATACAGTGGGAGAGATGGGTTAAATCAAGTAAAGAACAGCAAAACGCAAAATA 6706  
1320 CTTGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1379  
6707 CTTGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6766  
1380 AACTGGAAGTGTGCATCACACAGTGTCTCATCCAGCTGGCAATTTGAGCAACAAAGAAAGT 1439  
6767 AACTGGAAGTGTGCATCACACAGTGTCTCATCCAGCTGGCAATTTGAGCAACAAAGAAAGT 6826  
1440 GACACCCACTGCGAGTACTCAGCATGATGATGATGATGATGATGATGATGATGATGATGATG 1499  
6827 GACACCCACTGCGAGTACTCAGCATGATGATGATGATGATGATGATGATGATGATGATGATG 6886  
1500 AATATGTCATAGACAGATGCTGTGTCAGCATGATGATGATGATGATGATGATGATGATGATG 1559

6887 AATATGTCATAGACAGATGCTGTGTCAGCATGATGATGATGATGATGATGATGATGATGATG 6946  
1560 ATCAAAAGGGGAAACACAGCGGTTGAGTAAAGGGGAAACACATTCAGTCTAGTACCA 1619  
6947 ATCAAAAGGGGAAACACAGCGGTTGAGTAAAGGGGAAACACATTCAGTCTAGTACCA 7006  
1620 ATATGTTGGTGAAGAACACAGCAATGCTGAGAACATACATGATCTCAGCGCAGAGATCA 1679  
7007 ATATGTTGGTGAAGAACACAGCAATGCTGAGAACATACATGATCTCAGCGCAGAGATCA 7066  
1680 ATGCGAGATGGAACCGAAACTCTGTTCTGAGTCACTCGGCA----- 1723  
7067 ATGCGAGATGGAACCGAAACTCTGTTCTGAGTCACTCGGCAAGGTAAGATTTTGTG 7126  
1724 ----- 1723  
7127 AATCATGAGGAATTTTTCCTTTTAAATGACTGAATCAACATTTATCTGTATGAGGAA 7186  
1724 ----- 1723  
7187 TAATATTGGTGATAAACAAATGTTAAGAAATATGATACAAATGTTTATTTATATGCTTTCC 7246  
1724 ----- 1723  
7247 ACTGTTCTTCTTACTTATGTTTGTGATCTCTTTTGTGTCGTGATGTCATGTCATGT 7306  
1724 ----- 1723  
7307 GT 7366  
1724 -----AGGTTT 1729  
7367 AGACTCATATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7426  
1730 CTTGAGCTGAGCATGATATCCAAATTTATGTCTGACCTTTCATGAGCAGAGTCTAGCCAGA 1789  
7427 CTTGAGCTGAGCATGATATCCAAATTTATGTCTGACCTTTCATGAGCAGAGTCTAGCCAGA 7486  
1790 AGAAAAAGAACAAACCTTGAGTCTGTGAAACACAGCCATGATGATGATGATGATGATGATG 1849  
7487 AGAAAAAGAACAAACCTTGAGTCTGTGAAACACAGCCATGATGATGATGATGATGATGATG 7546  
1850 CCATGGATATTTGTTGAACTGCTAGCTAATAAACACAGATGAGAGGCGCTTATGATCTGAGA 1909  
7547 CCATGGATATTTGTTGAACTGCTAGCTAATAAACACAGATGAGAGGCGCTTATGATCTGAGA 7606  
1910 CTGATTTCTGACATCAACCGTATTCATTCGAAAGACAACTGCTGTGATGATGATGATGATG 1969  
7607 CTGATTTCTGACATCAACCGTATTCATTCGAAAGACAACTGCTGTGATGATGATGATGATG 7666  
1970 TAGTAGCTGCCAAGATGTTTCAGATTTATGATCAAGTGTGTTTGAACACTAATTTCCCAAC 2029  
7667 TAGTAGCTGCCAAGATGTTTCAGATTTATGATCAAGTGTGTTTGAACACTAATTTCCCAAC 7726  
2030 AGAAGTCTTTGGCATCCCAAGTACAGAAAGGATTAACGGGTCTATTTGGCATTTGACCA 2089  
7727 AGAAGTCTTTGGCATCCCAAGTACAGAAAGGATTAACGGGTCTATTTGGCATTTGACCA 7786  
2090 CACAGAGTCTCCACATCTCCAGAACTTTCAGTCTCTCTCAGAAACAGCAGACACATTTGC 2149  
7787 CACAGAGTCTCCACATCTCCAGAACTTTCAGTCTCTCTCAGAAACAGCAGACACATTTGC 7846  
2150 GGATGGAAGAAATGGTCACTATTCTGCAAGCTCACCCTATTTTTCATCATCATGATGATC 2209  
7847 GGATGGAAGAAATGGTCACTATTCTGCAAGCTCACCCTATTTTTCATCATCATGATGATC 7906  
2210 AGTATATTGCTGAAGCACCACCTGACATTTGGGCGCGTAAAGACGCAAGAGCTTAAAGT 2269  
7907 AGTATATTGCTGAAGCACCACCTGACATTTGGGCGCGTAAAGACGCAAGAGCTTAAAGT 7966  
2270 GGGAGCAATTTAAGGCCACTACAAGAAATTTCTCCAGCAGCAACATGTTGGTGTCTCAATTTA 2329

Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPs; inheritance; characteristic; growth; development; disease resistance; environmental adaptability; quality; yield; molecular marker; multigene trait; plant breeding; corn tassel; gene; ss.

Zea mays.

US2001051335-A1.

13-DEC-2001.

16-APR-1999; 99US-00294093.

21-APR-1998; 98US-0082567P.

(LALG/) LALGUDI R V.

(ITOL/) ITO L Y.

(SHER/) SHERMAN B K.

Lalgudi RV, Ito LY, Sherman BK;

WPI; 2002-163647/21.

Novel purified corn tassel-derived polynucleotide useful for determining altered gene expression, to recover regulatory elements and to follow inheritance of desirable characteristics through hybrid breeding programs.

Claim 1; SEQ ID NO 2759; 201pp; English.

The present sequence describes a purified corn tassel-derived polynucleotide sequence (cdps) comprising a nucleic acid sequence selected from those given in ABL70627 to ABL76833. The cdps sequences encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I) can be used for determining altered gene expression, to recover regulatory elements and to follow inheritance of desirable characteristics through hybrid breeding programs. (I) are also useful in the evaluation, and alteration of desired characteristics associated with growth and development, disease resistance, environmental adaptability, quality and yield, and as molecular markers for studying inheritance of multigene traits in a plant breeding program. (I) can be used to produce a tassel-specific profile of gene transcription, a transcript image, to clone regulatory elements for use in transformation vectors, to express a polypeptide, to identify, isolate or extend identical or related corn tassel nucleic acid sequences from DNA libraries, in nucleic acid hybridisation or amplification technologies, as query sequences to determine homology of known sequences, as probe for use in Southern or Northern hybridisation, and to identify the presence of and/or to determine the degree of similarity between two (or more) nucleic acid sequences

Sequence 279 BP; 70 A; 74 C; 73 G; 56 T; 0 U; 6 Other;

Query Match 2.3%; Score 73.4; DB 6; Length 279;

Best Local Similarity 71.3%; Pred. No. 2.6e-10;

Matches 122; Conservative 0; Mismatches 45; Indels 4; Gaps 2;

2720 ATGGAAGATTCCATTGACATTGGAAGACTTATACGGCATCA---GCTGATGATCTGC 2776

105 ATGGAAGCCACCGCTGANACTGGAAGACTTCTCGGGCTCATTTCCAGCAAGACTTGC 164

2777 ACAGACCTTTTACGCCCATCTAGAGTTGGTGTGCTTGGCTCTCCCTGCTCCACAAGGAAA 2836

165 GGAGGCCCTTTAGCNCCTACCCTCGANTCGGTGTGCTCGGTTCAATTCGTGCACAGAGA 224

2837 TTGCAAACTGGTCGGAGAACTGTGGCACACAATCTGTGGTTATAGTTAGGAG 2887

225 TCGCAAACTGGTCTGGAA-TSCGGGCCGACGCTCTGGGTANAGACTAGGTG 274

RESULT 5

ADAT71938/c

ID ADAT71938 standard; DNA; 2000 BP.



CC does not express a gene associated with response to pathogenic infection,  
or in a corresponding incompatible or compatible interaction. (MI) is  
useful for conferring resistance to resistance or tolerance to a plant to  
bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
xx  
Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other:  
SQ

Query Match 1.6%; Score 52; DB 7; Length 2000;  
Best Local Similarity 10.0%; Pred. No. 0.0012;  
Matches 60; Conservative 273; Mismatches 268; Indels 1; Gaps 1;

Qy	280	AGTTCACGCCCACTTTCTGTAGCAAAAGTTTCGACGATGGGATTCGTGCAAGTCGCTTGAT	333
Db	316	RYMSRMAMVTKXMYTYGYKMKRGWAGRMWRSYCRWMSKACYTYNWRWMTTRRRW	375
Qy	340	AAAGTTGAAAACTTCAGATAATGGAACACGACCAAGAAGACTCTCCGCCAAAGACAGAAATGCG	399
Db	376	AKKSERTSRKKRWCMERKRYKSMRGYSNRMSCKGARNMKRCSRGRAMWKGCRGOMTCR	435
Qy	400	ACAAGTGATGGTTGCTCCATCACATTTGTTGGGAGCACITTTCTGGCTCTGATGCTTGCT	459
Db	436	MKSYGMMRWKSWKRNASKYKMSRMYSRWRKKKSRTTMMGKTGGWNGTGRGCRFKKSG	495
Qy	460	TCCCAAAAAGTGTTCTCTAGCACACAATCATCTCAAGGGAAGAATCTGATAGATCAACT	519
Db	496	MKRCRRRRWRGMYRMRWRYMSARYTMRYSACRKYSYSAARKARCYRGRKGYYNAGMW	555
Qy	520	CTTCCAAAGAGTGTGCAAGAGGAAGCAATGACTCCAAATGCAATGGCGCTTCTCGCAAGAAAT	579
Db	556	MKRYRMYRYMMWYRYRKYSCSWYCKMSYYASCHKSARKAKMCKRSMASKSXVR	615
Qy	580	GGAGTGTGAGGCCAATACTGATTCACCAATGAAGATTGTCAGAGGGCCAGCCCAAAAT	639
Db	616	SRKCRKCAKSSASAKRYMMMGMTSGSRMSRWKSYCTCYWRKWSMKSTCTTWMYMSKYT	675
Qy	640	TATGATGTGGCAGCAAAATGTCTCTGAGACAACACACTTCTGTCATGTTGGGCGCTTA	699
Db	676	YAKYGSYWRYYRACWYMWRYRYRYRYSYNTYMAWYTSRMAKTGMKYSGRYWTSWVK	735
Qy	700	GAAGTTCGCCAGATTACATGGCATGAGTAAGTAATGGTGCAGATCAAACCTCCATCCACT	759
Db	736	YCKCSWKYSRMMYYMSWNAKTWQWERYATRWMMWYRYSMKWYTWCTWMSGYTWIWTAT	795
Qy	760	CCAAAACTTCTTG-AAAGTGGTCTCAAAAGAAATGAGATGAAAAATCGAAAAACTGAAGA	818
Db	796	YMKRYMYWTKTYTYWYSATYTWGTWAAWMAWNAKTQWRMGMTCAKTRGARAKARYWNAWA	855
Qy	819	GACTCTTGTTGCTGAGCAGTGCATTTGACCAAGATCCTTAACCCATCTCTCGGAAGGA	878
Db	856	TWATCRWTKGKAARWATWMAKARRYMSWRAWYTYTTRTRYTKTCWAKRAGWSW	915
Qy	879	AC 880	
Db	916	AY 917	

RESULT 7  
ABX85499  
ID ABX85499 standard; cDNA; 274 BP.  
XX  
XX  
AC ABX85499;  
XX  
DT 24-APR-2003 (first entry)  
XX  
DE Corn ear-derived polynucleotide (cpd) #3959.

XX Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022; SATMON023;  
 KW structural gene; functional gene; regulatory gene;  
 KW corn ear-specific profile; gene transcription; gene expression;  
 KW hybrid plant; desirable trait expression; plant breeding program;  
 KW inheritance; desired characteristic; growth; development;  
 KW disease resistance; environmental adaptability; quality; yield;  
 KW

KW	multigene trait; plant; gene; ss.
XX	
XX	Zea mays.
XX	
XX	US6476212-B1.
XX	
XX	05-NOV-2002.
PD	
XX	
XX	14-MAY-1995; 99US-00313294.
PF	
XX	
XX	26-MAY-1998; 98US-0086722P.
XX	
XX	(INCY-) INCYTE GENOMICS INC.
PA	
XX	
XX	Lalgudi RV, Ito LY, Sherman BK;
PI	
XX	
XX	WPI; 2003-208840/20.
DR	
XX	
XX	Novel purified corn-ear derived polynucleotide useful as hybridization
PT	probe for detecting polynucleotide in sample, and for identifying,
PT	evaluating, and altering desired characteristics associated with growth,
PT	development.
PT	
XX	Example: SEQ ID NO 3959; 390bp; English.
PS	

The present invention relates to the isolation of corn ear-derived polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022 and SATMON023. Some of the cdps uniquely identify structural, functional, and regulatory genes of corn ear. The polynucleotides sequences are useful for detecting cdps in a sample, for producing a corn ear-specific profile of gene transcription, for detecting altered gene expression in inbred or hybrid plants, and for screening several molecules for specific binding to the polynucleotides. The cdps are useful to identify, isolate, or extend identical or related corn-ear nucleic acid sequences from DNA libraries, and in nucleic acid amplification or hybridisation techniques to follow the expression of desirable traits through plant breeding programs. Preferably, the cdps are used to identify, evaluate, alter, or follow the inheritance of desired characteristics associated with growth and development, disease resistance, environmental adaptability, quality, and yield of corn. The cdps are also useful as molecular markers for studying inheritance and multigene traits in a plant breeding program. The cdps are useful for producing purified corn-ear polypeptides by recombinant techniques. They are also useful in diagnostic assays to detect or confirm conditions or diseases associated with abnormal levels of cdp expression. AX81541-AX89140 represent corn ear-derived polynucleotides (cdps) of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO web site at [seqdata.uspto.gov/psipdsIDENTry.html](http://seqdata.uspto.gov/psipdsIDENTry.html)

Sequence 274 BP: 67 A: 75 C: 75 G: 56 T: 0 U: 1 Other;

Query Match	1.4%	Score 45.4;	DB 7;	Length 274;
Best Local Similarity	68.2%;	Pred. No. 0.029;		
Matches 107;	Conservative	0;	Mismatches 46;	Indels 4;
			Gaps 3;	

QY	2720	ATGGAAGAAGATTCAATTGACATTCGAGAGACTTATACCGGATCATGCTGCATGA--TCTGCA	2777
Db	105	ATGGAAGCCAAACGGCTGACATCGAAGACTTGTCTCGGGCTCAATTTCCAGCAAGACTTGC	164
QY	2778	CAGACCTTTTACGCCCCACATCCCTAGAGTTGGTGTGCTTGCTCCTTGGTCGAGAAGGAAT	2837
Db	165	GCAGCGCTTTTACGCCCTTCAACCCCTCGTTCGGTGTGCTCGGTTCAT--TGTGTCAGCAGGAGAT	223
QY	2838	TGCAAACTCGTTCGGAGAACTGTGGGCACAAATCTGGT	2874
Db	224	CGCAAA--TGTGTTCGAGAAATTCGCGGCCGCGCATCTCGGT	259

RESULT 8  
AAC44195  
ID AAC44195 standard; DNA; 574 BP.  
XX

[illegible]

	Query Match	1.3%; Score 42.8; DB 3; Length 574;	
Best Local Similarity	57.5%;	Pred. No. 0.25;	
Matches	77; Conservative	0; Mismatches	57; Indels
		0; Gaps	0;
QY	1011	GA CGGATAAAGAAGTTGATGAAGACGACGACGACACAGCAAGAAACCGCACTGCCCGAGCTGA	1070
DB	122	GAAGTTGAAGAAGTTGATGAAGAAAGAAGAAAGAAAGAGGAAATGTGTGATGATGTTGA	181
QY	1071	TGTTTCAGATGCAAAAGCTTTGCGGAGAAAGCCAAAAAGGTGCGGCTTCTATCAGAAAT	1130
DB	182	TGATGATGATGCTGGAAATGCTCGGAAAAAAGCAAGAAAGATGAAGATTTCACATCACT	241
QY	1131	TATTAATGCTAACCC	1144
DB	242	TTCAGACAAATATC	255

RESULT 9  
ABQ69245.18  
Continuation (19 of 31) of ABQ69245 from base 1800001 (*Listeria innocua* DNA sequence #58

Query Match	1.3%;	Score 42.2;	DB 6;	Length 110000;
Best Local Similarity	47.0%;	Pred. No. 7.8;		
Matches 131;	Conservative 0;	Mismatches 148;	Indels 0;	Gaps 0;
574	AAGAATGGAGCTGCTCAGCGCCAATACGTATTCACCAATGAAGAATTTCGAAGCGCCAGCC	633		
27998	ACAAAAAAGGACATGAATTTATGGCCAATATCCCAACAGAGAAGTTTTCTGTGTGGC	28057		
634	CAAAATTATGATGTGGCAGCAATGTCTCTGAGGACAAACACTTCGTGTGATGTGGGGCT	693		
28058	GATAAATCTAAAGTGAAGGTTAATGTTTCAAGCAGCAACCACTTAGTTATGCGGAAT	28117		
694	TTACTTGAAGTTTCCCAGAGATTACATGGCCACATAGAAGTAAATGTGTCAGATCAACTTCCA	753		
28118	ATTATTGATGACTTCAAAATCACTTTCGAAAAAGGTCGTATTGTGTGGCGTTGAAGCTGCA	28177		
754	TCCACTCCAAACTTCTCGAGTGGTCCCTCAAGAAGAAATCAAGTGAATAATCGAATAACT	813		
28178	TCUGGCGAAGAAATTTTAAAGATTTTAACTCGCAGCGTGAAGGTTCTCAITATTAGGC	28237		
814	GAAGAGACTCTTTGTGCTCAGCAGTGCAAATTTGACCAA	852		
28238	GAAGTGGCTCTAGTTCAGATCATCATCCCTTAITTCCAA	28276		

RESULT 10  
ABQ67194/c  
ID ABQ67194 standard; DNA; 319630 BP.  
XX  
XX ABQ67194;  
AC  
XX  
XX  
XX 29-AUG-2002 (first entry)  
XX  
XX Listeria innocua contig DNA sequence #7.  
XX  
XX Antibacterial; Listeria; food contamination; mutational analysis;  
KW infection; ds.  
KW  
XX  
XX Listeria innocua.  
OS

XX PN WO200228891-A2.  
XX PD 11-APR-2002.  
XX PF 04-OCT-2001; 2001WO-FR003061.  
XX PR 04-OCT-2000; 2000FR-00012697.  
XX PA (INSP ) INST PASTEUR.  
XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX PI Kunst F, Glaser P;  
XX PI WPI; 2002-332479/37.  
XX DR New genomic sequences from *Listeria* species, useful for detection,  
PT treatment and prevention of infection, also related polypeptides,  
PT antibodies and modulators.  
XX PS Claim 5; SEQ ID NO 7; 180pp; French.  
XX CC The present invention relates to nucleic acid sequences (AB067188-  
CC ABQ71212) from *Listeria* sp. The sequences are useful as probes and  
CC primers for identification and/or detection of *Listeria* (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of gene  
CC expression. Proteins encoded by the nucleic acid sequences can be used to  
CC screen for compounds that modulate gene expression, replication and  
CC pathogenicity of *Listeria* (potential therapeutic agents), also for  
CC treating infections by *Listeria*, and are useful as immunogens in anti-  
CC *Listeria* vaccines. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 319630 BP; 105207 A; 55428 C; 66726 G; 92263 T; 0 U; 6 Other;  
Query Match 1.3%; Score 42.2; DB 6; Length 319630;  
Best Local Similarity 47.0%; Pred. No. 14;  
Matches 131; Conservative 0; Mismatches 148; Indels 0; Gaps 0;  
QY 574 AAGATGGAGCTGCTGAGGCCAATCTGATTCCCAATGAAGAATTTGCAAGGGCCAGCC 633  
Db 289425 AACAAAAAGGACATGATTTATGGCCATATGCGCAACAGAGAATTTCTGTGTGCG 289366  
QY 634 CAAATATATGTCGACCAATGCTCTGAGGACACACTCTGTGATCTTGGGCT 693  
Db 289365 GATAAATCTAAAGTGAAGTTATGTTTCAAGCAGCAACCACTTAGTTATGCAGAAT 289306  
QY 694 TTACTTGAAGTTCCCGAGATTACATGCGCACATAGAGTAATGGTGACATCAACTCCA 753  
Db 289305 ATTATTGATGACTTCAAAATCACTTTTGGAAAAGGTGGTATTTGCGGTGAAGCTGCA 289246  
QY 754 TCCACTCCAAACTTCTGAGTGGTCTTCAAAAGAAATGAGATGAATGGAATGGAAT 813  
Db 289245 TCTGCGAAGAAATTTAAAGATTTTAAAGATTTTAAAGATTTTAAAGATTTTAAAG 289186  
QY 814 GAAGAGACTCTTGTCTGTCGAGTGCATTTGACCCAAA 852  
Db 289185 GAAGTGGCTTAGTTCAGATCCATCCCTATTTCCTCA 289147  
RESULT 11  
AA590679  
ID AAS90679 standard; cDNA; 246 BP.  
XX AC AAS90679;  
XX AC AAS90679;  
XX DT 13-FEB-2002 (first entry)  
XX DE DNA encoding novel human diagnostic protein #26483.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PR 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX PI WPI; 2001-639362/73.  
XX DR P-FSDB; ABG26492.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX PS Claim 1; SEQ ID NO 26483; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical disorders  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 246 BP; 126 A; 20 C; 77 G; 23 T; 0 U; 0 Other;  
Query Match 1.3%; Score 41.4; DB 5; Length 246;  
Best Local Similarity 60.0%; Pred. No. 0.39;  
Matches 69; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 938 AGAATGTGACGACATCTGCATGAGCCATCTGAGAGTGTGTTCTCAAGAGAGCTCCA 997  
Db 71 AGGACTGTGAAG 130  
QY 998 AATCTAAG 1052  
Db 131 AG 185  
RESULT 12  
ABN90734  
ID ABN90734 standard; DNA; 1269 BP.  
XX AC ABN90734;  
XX AC ABN90734;  
XX DT 24-JUL-2002 (first entry)  
XX DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:197.  
XX KW







QY	473	CTCCTAGCACACAATCATCTCAAGGGAAGAATGCTGATAGATCAACTCTTCCAAAGAGTG	532
Db	1052	CTATAAGTGAAGCAATCGAAGCTAAGCAAGGAGGCGGTGATTACGTTCAATTATAG	1111
QY	533	TGCAAGAAGGCAATGACTCCAAATGCAATCGGCTTCTGSCAAGAATGGAGCTGCTGAGG	592
Db	1112	AAGAAGCATTAATGACATCATGTATGATGTACCTTCTTGTGAAATGTAAGTAAAGTAG	1171
QY	593	CCAATACTGATTCAACCAATGAAGAATTTGCAAGGCGCCAGCCCAAAATTTATGATGTGGCAG	652
Db	1172	TCATTACTGACAAACAATTAATGAGAAATTCAGGCTGATATATGATGATGAGGGA	1231
QY	653	CAAAATCTCTTGAGGACACACTTCTGTT	681
Db	1232	ATCTAATTAATAGAACAGACATCTGCT	1260

Search completed: May 16, 2004, 20:12:21  
Job time : 1204 secs

OM nucleic - nucleic search, using sw model

Run on: May 16, 2004, 19:21:07 ; Search time 219 Seconds  
(without alignments)  
8035.385 Million cell updates/sec

Title: US-09-721-114-1\_COPY\_655\_3825  
Perfect score: 3171  
Sequence: 1 atggagattgttcagtaga.....acgagtatattgattaccgc 3171

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 692709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	62.8	2.0	7218	1	US-08-232-463-14
2	45.4	1.4	274	4	Sequence 14, Appli
3	41.4	1.3	1269	4	Sequence 3959, Ap
4	40.6	1.3	297	4	Sequence 197, App
5	40.4	1.3	686	4	Sequence 215, Appl
6	39	1.2	832	4	Sequence 9, Appli
7	38.2	1.2	335	4	Sequence 2813, Ap
8	38	1.2	639	4	Sequence 45, Appl
9	37.4	1.2	19619	4	Sequence 809, App
10	37.4	1.2	508	4	Sequence 14, Appl
11	37.4	1.2	1407	4	Sequence 535, App
12	36.8	1.2	576	4	Sequence 2195, Ap
13	36.8	1.2	8648	4	Sequence 6914, Ap
14	36.8	1.2	17341	4	Sequence 2, Appli
15	36.4	1.1	2277	1	Sequence 1, Appli
16	36.4	1.1	2277	1	Sequence 2, Appli
17	36.4	1.1	2277	1	Sequence 2, Appli
18	36.2	1.1	1140	3	Sequence 32, Appl
19	36.2	1.1	1140	3	Sequence 32, Appl
20	36.2	1.1	1525	2	Sequence 1, Appli
21	36.2	1.1	1525	2	Sequence 3, Appli
22	36.2	1.1	1526	3	Sequence 3, Appli
23	36.2	1.1	1674	2	Sequence 1, Appli
24	36.2	1.1	1674	3	Sequence 1, Appli
25	36.2	1.1	1674	4	Sequence 1, Appli
26	36.2	1.1	1674	5	Sequence 1, Appli
27	36.2	1.1	1937	3	Sequence 7, Appli

28	36.2	1.1	1997	3	US-08-510-133A-34	Sequence 14, Appli
29	36.2	1.1	1997	4	US-09-355-700-7	Sequence 7, Appli
30	36.2	1.1	1997	4	US-08-601-132-32	Sequence 32, Appli
31	36.2	1.1	1997	4	US-08-671-573B-32	Sequence 32, Appli
32	36.2	1.1	2031	4	US-08-706-054A-1	Sequence 1, Appli
33	36.2	1.1	2031	4	US-08-706-054A-2	Sequence 2, Appli
34	36.2	1.1	2031	4	US-09-313-299-1	Sequence 1, Appli
35	36.2	1.1	2031	4	US-09-313-299-2	Sequence 2, Appli
36	35.8	1.1	289	3	US-09-007-005-17	Sequence 17, Appli
37	35.8	1.1	289	3	US-09-244-796-17	Sequence 17, Appli
38	35.8	1.1	489	4	US-09-134-001C-2451	Sequence 2451, Ap
39	35.8	1.1	2673	4	US-09-519-232-73	Sequence 73, Appli
40	35.8	1.1	11283	2	US-08-603-753D-3	Sequence 3, Appli
41	35.8	1.1	11283	3	US-09-099-753-3	Sequence 3, Appli
42	35.8	1.1	11283	3	US-08-986-106-3	Sequence 3, Appli
43	35.8	1.1	11385	2	US-08-639-501-1	Sequence 1, Appli
44	35.8	1.1	11385	3	US-09-044-946-1	Sequence 1, Appli
45	35.8	1.1	11385	3	US-09-044-908-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: PALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/232,463  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 INMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgt-F15  
; US-08-232-463-14

Query Match 2.0%; Score 62.8; DB 1; Length 7218;

	Query Match	1.4%	Score 45.4;	DB 4;	Length 274;
	Best Local Similarity	68.2%;	Pred. NO. 0.00057;		
	Matches 107;	Conservative	0;	Mismatches 46;	Indels 4; Gaps 3;
QY	2720	ATGGAAGAGATTCCATTGACATTCCAGAGATTATCAAGGCATACAGCTGCATGA--TTTGCA	2777		
DB	105	ATGGAAGCCAAACGGCTGCATCTGGAAGACTTGTCTCGGGGTCTATTTCCAGCAGAGATTGC	164		
QY	2778	CAGACCTTTTACGCCCATCTCTAGAGTTGGTGTGCTTGGCTCCCTTCTGCAGAAAGAAAT	2837		
DB	165	GCAGCCCTTTACGCCCTCACCCCTCGTTCGGTGTGCTCGGTTCAT--TGCCTGCAGCAGAGAT	223		

```

RESULT 4
US-09-134-001C-215/c
; Sequence 215, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 215
; LENGTH: 297
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-215

```

Query Match 1.3%; Score 40.6; DB 4; Length 297;  
Best Local Similarity 49.8%; Pred. No. 0.02;



```

Query Match          1.2%; Score 38; DB 4; Length 639;
Best Local Similarity 50.5%; Pred. No. 0.23;
Matches 92; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY      1100 AGCCAAAAAAGGTGGCGGCTTCTATCAGAAAATTATAAATGCTAAACCAAGTTTGAGGATTCTA 1159
Db      53  AACCAAAAAAAGGTGAATATTATGGCAGCAATATTATTAAATCAAAACCAATTCACAGATTTCAA 112
QY      1160 GAAGTCAGCAAGCTTCATCTGGAATATCGCGCTGATCCCTGTGAGGATGATAGAGTACCA 1219
Db      113  CCCAAGCCTATGTGAATGTGTAGTTTAAACCATCATCACTTCTGATGATGTTAAAGCGAGCT 172
QY      1220 TCCCGGTCCCGATGGAAGTAAGCATGGATATTCCTGTGTAGCAACCATACAGTGGGAGAAG 1279
Db      173  GGGCGATTTCATGTGTTTATCCACATGATTTTACTTTTGTGGCCCAACAGAGCTAGAGG 232
QY      1280 AT 1281
Db      233 AT 234

RESULT 9
US-09-596-002-14
; Sequence 14, Application US/09596002
; Patent No 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 14
; LENGTH: 19619
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURES:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte template ID No. 6632636 14
; PUBLICATION INFORMATION:
; US-09-596-002-14

```



Db 253 AGCAATGGGATAATAACCTTGGGATCAACTTGGTGATTAATCTCTCTGAAA 203

RESULT 11

US-09-328-352-2195

Sequence 2195, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

FILE REFERENCE: GTC99-03PA

CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 2195

TYPE: DNA

LENGTH: 1407

ORGANISM: Acinetobacter baumannii

US-09-328-352-2195

Query Match 1.2%; Score 37.4; DB 4; Length 1407;

Best Local Similarity 43.6%; Pred. No. 0.63;

Matches 167; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 489 ATCTCAAGGGAAGAATGCTGATAGATCAACTCTTCCAAAGAGTGTGCAAGAAGGCAATGA 548

Db 558 AGCTAAACGTAAAGCTGAGACTGAAGCAAGACTATAAAGCTGAAGCTCAAAAAGCTGAAA 617

QY 549 CTCCTAAATGCAATGCGCTTCTGGCAAGATGAGCTGCTGAGGCCAATGACTGATCACC 608

Db 618 CGCTAAACTTCAAGCACACAAGAAAGCTAAACGAAAAGCCGAGCAGATGCTTAAAGCCAA 677

QY 609 AATGAAAGATTTTCAAGGCCAGCCCAAAATTTATGATGGCGAGCAAAATGCTCTCAGGA 668

Db 678 ACAGCAAAAGCTGCGAAGAGCCCAAGCTTAAGCTGAATCAGATGCTAAGGCTAAGCA 737

QY 669 CAACACTTCTGTGTGATGTTGGGGCTTTACCTGAAGTTCCCGAGATTACATGGCACATAGA 728

Db 738 ACAAGCAGCTGATAACCGGAAACGTAAGCCGAGCCGATGCGAAAGCTTAAACAGCAAAA 797

QY 729 AGTAAATGTCAGATCACTCCATCCATCTCCAAACTTTCTGAAGTGTCTCTCAAAAG 788

Db 798 AGAGCTGAAGATGCTTAAGCTTAAGCTTAAGCTTAAGCTTAAGCTTAAGCTTAAGCTTAAG 857

QY 789 AAATGAAGATGAAATGGAAGAACTGAAGAGAGACTCTTGTGCTGAGCAGTGCATTTGAC 848

Db 858 AGCTGAAGATGCTTAAGCGCAAGCTGAGCGCGATGCGAAAGCTTAAACAGCAAAAAGCGC 917

QY 849 CAAAGATCTTACCCCAATGCTG 871

Db 918 TGAAGATGCTTAAACGCAAAAGCTG 940

RESULT 12

US-09-489-039A-6914

Sequence 6914, Application US/09489039A

Patent No. 6610836

GENERAL INFORMATION:

APPLICANT: Gary Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

FILE REFERENCE: 2709.2004001

CURRENT APPLICATION NUMBER: US/09/489,039A

CURRENT FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: US 60/117,747

PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 14342

SEQ ID NO 6914

LENGTH: 576

TYPE: DNA

ORGANISM: Klebsiella pneumoniae

US-09-489-039A-6914

APPLICANT: Ross, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESS: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Monroy, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

TELEX: 706141

SEQUENCE CHARACTERISTICS:

LENGTH: 508 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: UNKNOWN

ORIGINAL SOURCE:

ORGANISM: PORPHYROMONAS GINGIVALIS

FEATURE:

NAME/KEY: misc feature

LOCATION: 1...508

US-09-221-017B-535

Query Match 1.2%; Score 37.4; DB 4; Length 508;

Best Local Similarity 47.6%; Pred. No. 0.29;

Matches 110; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 758 CTCCAAACTTTCTGAAGTGGTCTCTCAAAAGAAATGAAGATGAAATGGAAGAACTGAAG 817

Db 433 CGCCACAGCATTCGAAAGATTTCTTGGCTATGATCTCTCCACAGATTTGGAAG 374

QY 818 AGACTCTTGTGTCAGCAGTGCATTTGACCAAGATCTTACCCCAATGCTTGGAAAG 877

Db 373 ATCAGCAGGATGATATGAAGCCGCAAAATGATACCGATGATCTAAGGAAGACGCAAC 314

QY 878 AACCTGATCAGGTTGTCAGCAGTGCATTTGACCAAGATCTTACCCCAATGCTTGGGC 937

Db 313 ACCTTACCACTTCGCTGATGAGGCGCAAAATGATACCGATGATCTAAGGAAGATCCGAGATATGATGGC 254

QY 938 AGAAATGTGAGCAGATCTGCAATGAGCCATGTCAGAGATTTGTTCTCAAAA 988

```

; LOCATION: (5046)..(6307)
; OTHER INFORMATION: exon 6
; NAME/KEY: intron
; LOCATION: (6308)..(6447)
; NAME/KEY: exon
; LOCATION: (6448)..(8065)
; OTHER INFORMATION: exon 7
; NAME/KEY: intron
; LOCATION: (8066)..(8300)
; NAME/KEY: exon
; LOCATION: (8301)..(8648)
; OTHER INFORMATION: exon 8
; US-09-415-946-2

```

Query Match	1.2%;	Score 35.8;	DB 4;	Length 8648;
Best Local Similarity	50.6%;	Pred. No. 3.7;		
Matches 89;	Conservative	0;	Mismatches 87;	Indels 0;
Gaps	0;			
1724	AGGTTTC	CGAGCTG	GAGCATG	ATCCAAATTATGCTCGACCTTCATGAGCAGAGTCTAC 1783
6446	AGGATGCTTCTGGT	CGAGATGAAAGGGGAG	AGCTGTTATGGTCC	AGACATCTGCGAG 6505
1784	CCACAGACAAAAGAAAG	CAAAAACCTTGAGTGC	ACTCGTGCAAAAA	CAGACCATGATAGATG 1843
6506	CACCAAGAGGCCA	AAAGTCACGNTAG	AAAGGAGACTACGACTG	AAGGCCAAACACGATG 6565
1844	ACATCCCATCGATAT	TTTGACTGCTAGCT	TAAAAACACGATG	GAGAGGCGAGCTT 1899
6566	ATATTCCAAATGAGAT	GTGGAGCTCATGGC	CAAAAAACGATPAC	GAGAGGTGCTTT 6621

```

RESULT 14
US-09--415-946-1
; Sequence 1, Application US/09415946
; Patent No. 6376751
; GENERAL INFORMATION:
; APPLICANT: Sung, Z. Renee
; APPLICANT: Aubert, Dominique
; APPLICANT: Chen, Lingling
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Nucleic Acids That Control Reproductive Development in
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 018941-000110US
; CURRENT APPLICATION NUMBER: US/09/415,946
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,696
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 17341
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: genomic DNA (Ecotype Columbia) from CD82 clone
; NAME/KEY: promoter
; LOCATION: (1)..(3201)
; NAME/KEY: CDS
; LOCATION: join(4241..4335, 4448..4623, 4704..4823, 4903..4956,
; OTHER INFORMATION: EMBRYONIC FLOWER 1 (EMF1)
; NAME/KEY: exon
; LOCATION: (3202)..(3265)
; OTHER INFORMATION: exon 1
; NAME/KEY: intron
; LOCATION: (3266)..(4159)
; NAME/KEY: exon
; LOCATION: (4160)..(4335)
; OTHER INFORMATION: exon 2
; NAME/KEY: intron
; LOCATION: (4336)..(4447)
; NAME/KEY: exon
; LOCATION: (4448)..(4623)
; OTHER INFORMATION: exon 3

```

NAME/KEY: intron  
LOCATION: (4624)..(4703)  
NAME/KEY: exon  
LOCATION: (4704)..(4823)  
OTHER INFORMATION: exon 4  
NAME/KEY: intron  
LOCATION: (4824)..(4902)  
NAME/KEY: exon  
LOCATION: (4903)..(4956)  
OTHER INFORMATION: exon 5  
NAME/KEY: intron  
LOCATION: (4957)..(5045)  
NAME/KEY: exon  
LOCATION: (5046)..(6307)  
OTHER INFORMATION: exon 6  
NAME/KEY: intron  
LOCATION: (6308)..(6447)  
NAME/KEY: exon  
LOCATION: (6448)..(8065)  
OTHER INFORMATION: exon 7  
NAME/KEY: intron  
LOCATION: (8066)..(8300)  
NAME/KEY: exon  
LOCATION: (8301)..(8648)  
OTHER INFORMATION: exon 8  
US-09-415-946-1

Query Match 1.2%; Score 36.8; DB 4; Length 17341;  
Best Local Similarity 50.6%; Pred. No. 6.2;  
Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
QY 1724 AGTTTCTCCAGTGGAGCATGATATCCAAATTATGTCTGACCTTCATGAGCAGAGTCTAC 1783  
DB 6446 AGGATGCTTCTGGTGAGATAGAAAGGGGAGAGCTGTTATGGTCCAAAGAACATCATGGAG 6505  
QY 1784 CCAAG 1843  
DB 6506 CACCAAG 6565  
QY 1844 ACATCCCATGATATGTTGAACTGCTAGCTTAAACAGAGATGAGAGGAGCTT 1899  
DB 6566 ATATCCATGAGATAGTGGAGCTCATGCGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAG 6621

RESULT 15  
US-08-676-967-2  
Sequence 2, Application US/08676967  
Patent No. 5747317  
GENERAL INFORMATION:  
APPLICANT: COLLINS, KATHLEEN  
TITLE OF INVENTION: Human Telomerase  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Science & Technology Law Group  
STREET: 268 Bush Street, Suite 3200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,967  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UCS96-055  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 343-4341  
TELEFAX: (415) 343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-676-967-2  
Query Match 1.1%; Score 36.4; DB 1; Length 2277;  
Best Local Similarity 29.7%; Pred. No. 1.8;  
Matches 116; Conservative 66; Mismatches 206; Indels 3; Gaps 2;  
QY 668 ACAACACTTCTGTTGATGTTGGGGCTTTACCTGAGTTCCCGAGATTCCCGACATGCGACATAG 727  
DB 391 AARACNGTNTTYGNCARTTYGCGCTGTTGARGTNAAYATHCCNMGNAAARCCNGAY 450  
QY 728 AAGTAAATGGTGCAGATCAACCTCCATCCACTCCAAAACCTTCTTGAAGTGGTCCCTCAAAA 787  
DB 451 GGNAAATGTMGNGGNTTYGNTTGTNCARTTYAARAAAYTNTNGARGCGNGGNAARCCN 510  
QY 788 GAATGAAGATGAAAATGAAAACCTGAAGAGAC-TCTTGTGCTGAGCAGTGCATTTG 846  
DB 511 YTNARGGNATGAAYATGAARGATHAARGNMGNACNGTNGCTNGAYTGGCNGTN 570  
QY 847 ACCAAGATCCTAACCCCAATGCTTGGAAAGGAACGTGATCAGTTGCTGAGCAGTGCAT 906  
DB 571 GGNAAAGAYAAATYAARGAYACNCARWSNGTMSNGCNATHGGNGARGAARMSNCAY 630  
QY 907 TTGACCAAGATCCGAAACCAAGTCTCTGGGCGAGAAATGTGAGCAGATCTGCAATCGACCA 966  
DB 631 GARWSNAARCAICANGARWSGTNAARARGGNMGNARGARGARGAYATGARGAR 690  
QY 967 TGTGAAGAAGTTGTTCTCAAAAAGAGCTCCAAATCTTAAGAGAGAGACCGATAAGAGTTG 1026  
DB 691 GARGAARAYGAG--AYGAYGAYGAYGAYGAYGARGARGAYGNGTNTTYGAYGAYG 748  
QY 1027 ATGAAGAGCAGCAGCAGCAGCAAGCAAGCA 1057  
DB 749 ARGAYGARGARGARGAARAAAYATHGARNNA 779

Search completed: May 17, 2004, 01:46:55  
Job time : 224 secs

GenCore version S.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 16, 2004, 23:34:20 ; Search time 1288 Seconds  
(without alignments)  
11172.479 Million cell updates/sec

Title: US-09-721-114-1\_COPY\_655\_3825  
Perfect score: 3171  
Sequence: 1 atggagatgtgcagtaga.....acaggtatattgattaccgc 3171

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2947324 seqs, 2269024515 residues

Total number of hits satisfying chosen parameters: 5894648

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	3171	100.0	3896	9	US-09-828-068-1
2	3171	100.0	4310	15	US-10-447-135-1
3	2573.2	81.1	9455	15	US-10-447-135-3
4	168.8	5.3	871	13	US-10-425-114-4042
5	73.4	2.3	279	9	US-09-294-0938-2759
6	42.2	1.3	319630	16	US-10-398-221-7
7	42.2	1.3	3011208	16	US-10-398-221-2058
8	41	1.3	1260	13	US-10-282-122A-34801
9	40.6	1.3	5451	13	US-10-188-186-59
10	40.6	1.3	5761	13	US-10-302-172-23
11	40	1.3	412	10	US-09-918-995-22363
12	39.8	1.3	15122	15	US-10-311-455-748
13	39.8	1.3	3673778	15	US-10-312-841-2
14	39.2	1.2	1251	13	US-10-425-114-813

15	39.2	1.2	2522	9	US-09-745-763-114	Sequence 114, App
16	39.2	1.2	6065	15	US-10-311-455-477	Sequence 477, App
17	39.2	1.2	34475	16	US-10-085-117-1	Sequence 1, Appl
18	38.8	1.2	648	12	US-10-152-319A-2106	Sequence 2106, Ap
19	38.8	1.2	4725	12	US-10-041-018-147	Sequence 147, App
20	38.8	1.2	256525	13	US-10-087-192-451	Sequence 451, App
21	38.6	1.2	261	9	US-09-864-761-19524	Sequence 19524, A
22	38.6	1.2	261	9	US-09-864-761-25517	Sequence 25517, A
23	38.6	1.2	448	9	US-09-864-761-557	Sequence 557, App
24	38.6	1.2	513	9	US-09-864-761-8851	Sequence 8851, Ap
25	38.6	1.2	1042	13	US-10-027-632-324274	Sequence 324274,
26	38.6	1.2	1042	16	US-10-027-632-324274	Sequence 146, App
27	38.6	1.2	1062	9	US-09-770-445-146	Sequence 237, App
28	38.6	1.2	2001	16	US-10-310-154-237	Sequence 621, App
29	38.6	1.2	2100	9	US-09-938-842A-621	Sequence 621, App
30	38.6	1.2	2100	11	US-09-938-842A-621	Sequence 215, App
31	38.6	1.2	3282	17	US-10-149-310-215	Sequence 99680, A
32	38.4	1.2	2094	16	US-10-027-632-99680	Sequence 99680, A
33	38.4	1.2	2094	16	US-10-027-632-99680	Sequence 170, App
34	38.4	1.2	2916	9	US-09-925-302-170	Sequence 170, App
35	38.4	1.2	2916	13	US-09-925-302-170	Sequence 24, Appl
36	38.4	1.2	4139	16	US-10-439-388-24	Sequence 6666, Ap
37	38.4	1.2	7105	10	US-09-764-891-6666	Sequence 18143, A
38	38.2	1.2	298	9	US-09-864-761-18143	Sequence 45, Appl
39	38.2	1.2	395	9	US-09-894-844-45	Sequence 45, Appl
40	38.2	1.2	395	13	US-10-647-089-45	Sequence 45, Appl
41	38.2	1.2	395	16	US-10-388-902-45	Sequence 27, Appl
42	38.2	1.2	1617	15	US-10-197-666A-27	Sequence 82, Appl
43	38.2	1.2	1617	15	US-10-024-298A-82	Sequence 82, Appl
44	38.2	1.2	1617	15	US-10-042-211A-82	Sequence 82, Appl
45	38.2	1.2	1617	17	US-10-617-217A-82	Sequence 82, Appl

ALIGNMENTS

RESULT 1  
US-09-828-068-1  
; Sequence 1, Application US/09828068  
; Patent No. US20020157137A1  
; GENERAL INFORMATION:  
; APPLICANT: Moon, Yong-Hwan  
; APPLICANT: Chen, Lingling  
; APPLICANT: Sung, Zimay R.  
; TITLE OF INVENTION: NUCLEIC ACIDS THAT CONTROL REPRODUCTIVE DEVELOPMENT IN  
; TITLE OF INVENTION: PLANTS  
; FILE REFERENCE: 018941-001400US  
; CURRENT APPLICATION NUMBER: US/09/828,068  
; CURRENT FILING DATE: 2002-04-11  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 1  
; TYPE: DNA  
; LENGTH: 3896  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (241)...(3411)  
US-09-828-068-1

Query Match	100.0%	Score 3171;	DB 9;	Length 3896;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 3171;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	ATGGAGATTTGTCAGTAGATCAGGAGGAGCTGCTGTTGTTGGGACGAACTGTATGCTT	60	
DB	241	ATGGAGATTTGTCAGTAGATCAGGAGGAGCTGCTGTTGTTGGGACGAACTGTATGCTT	300	
QY	61	GCTCGTGTGGAACTGTGCTGTAGCCGACGTTGTTGGAGCTGACACGACGCTCGCTCAG	120	
DB	301	GCTCGTGTGGAACTGTGCTGTAGCCGACGTTGTTGGAGCTGACACGACGCTCGCTCAG	360	
QY	121	GATGACGCGCTGAAGCTGGTGTAGAGAACCGGACCAACCAATCGGAGCATTTCTCC	180	



QY	2341	CAAGCAGTTGCATCTGACTTCTACTCATGTGTCATGGGATCTTCCAGCAATTAAGCATCTCGC	2400
DB	2581	CAAGCAGTTGCATCTGACTTCTACTCATGTGTCATGGGATCTTCCAGCAATTAAGCATCTCGC	2540
QY	2401	CAACCAAGTAATTCGGCCACCTGGACCGCTATGCTGAAAGACGCGTTAACCAAGTCCATGCA	2460
DB	2641	CAACCAAGTAATTCGGCCACCTGGACCGCTATGCTGAAAGACGCGTTAACCAAGTCCATGCA	2700
QY	2461	AGAAAATTTTCCAAAGCACAAATAGCAACCATGGAAGCGAGTAAGTTATGTGATCGGAGAAAT	2520
DB	2701	AGAAAATTTTCCAAAGCACAAATAGCAACCATGGAAGCGAGTAAGTTATGTGATCGGAGAAAT	2760
QY	2521	GCTGGCAAGTAGTCTTGATCTCTTAAGATACCATGCTCGACGCGATCTTCTGAGAAAG	2580
DB	2761	GCTGGCAAGTAGTCTTGATCTCTTAAGATACCATGCTCGACGCGATCTTCTGAGAAAG	2820
QY	2581	ATGGATCCCATCAACATTAGCAAGCTTCCCAAATATGGAACCTTCTAGCAGGAACCCAGATG	2640
DB	2821	ATGGATCCCATCAACATTAGCAAGCTTCCCAAATATGGAACCTTCTAGCAGGAACCCAGATG	2880
QY	2641	GAGTCTCAACTTCATTAATCTCAGTAGTGACATAATCAGTACAAAGATCAACACGACCA	2700
DB	2881	GAGTCTCAACTTCATTAATCTCAGTAGTGACATAATCAGTACAAAGATCAACACGACCA	2940
QY	2701	TCATATGGCAGTAGTAACCTGGAATGAAAAGATGCCAATTGCAATTTCGAAGACTTATCACGGCAT	2760
DB	2941	TCATATGGCAGTAGTAACCTGGAATGAAAAGATGCCAATTTCGAAGACTTATCACGGCAT	3000
QY	2761	CAGCTGCAATGATCTGCAACAGACCTTTTAGCGCCACATCTCAGAGTTGGTGGCTTGCTGCTCC	2820
DB	3001	CAGCTGCAATGATCTGCAACAGACCTTTTAGCGCCACATCTCAGAGTTGGTGGCTTGCTGCTCC	3060
QY	2821	TTGCTGCAGAAAGGAATTTGCAAACTGGTTCGGAGAACTTGCGCACACAATCTGGTTATTAAG	2880
DB	3061	TTGCTGCAGAAAGGAATTTGCAAACTGGTTCGGAGAACTTGCGCACACAATCTGGTTATTAAG	3120
QY	2881	TTAGGAGTGTCAACAGGAATPAAATCCGATCAGATGAAACAGAAAGGAACATTTTGAAGCC	2940
DB	3121	TTAGGAGTGTCAACAGGAATPAAATCCGATCAGATGAAACAGAAAGGAACATTTTGAAGCC	3180
QY	2941	CTGAATCTCGAATGTTTTCCAGCAAAAATGGAATGCAATTGCAAGTTGGGTTCTGTTAGCTCC	3000
DB	3181	CTGAATCTCGAATGTTTTCCAGCAAAAATGGAATGCAATTGCAAGTTGGGTTCTGTTAGCTCC	3240
QY	3001	AGTGCAGATTTTTTATCAGCGAGGAACAGCATAGCTCAATCTTGGACCGAGGCAAGGGT	3060
DB	3241	AGTGCAGATTTTTTATCAGCGAGGAACAGCATAGCTCAATCTTGGACCGAGGCAAGGGT	3300
QY	3061	AAAATGGTTTCATCCCTTGGATCGGTTTGTGACAGCAGGATATCTGTAATCTAACAGAAC	3120
DB	3301	AAAATGGTTTCATCCCTTGGATCGGTTTGTGACAGCAGGATATCTGTAATCTAACAGAAC	3360
QY	3121	CCAGCTGATTTTACTACAATCAGTAAACGATAACGAGTATATGGATTACCGC	3171
DB	3361	CCAGCTGATTTTACTACAATCAGTAAACGATAACGAGTATATGGATTACCGC	3411

RESULT 2  
US-10-447-135-1  
; Sequence 1, Application US/10447135  
; Publication No. US2003019984A1  
; GENERAL INFORMATION:  
; APPLICANT: Hirochika, Hirochiko  
; APPLICANT: Yamazaki, Munee  
; APPLICANT: Miyao, Akio  
; TITLE OF INVENTION: A novel gene involved in brassinosteroid responses  
; FILE REFERENCE: MAFF-1 DIV  
; CURRENT APPLICATION NUMBER: US/10/447,135  
; CURRENT FILING DATE: 2003-05-27  
; PRIOR APPLICATION NUMBER: US 09/721,114  
; PRIOR FILING DATE: 2000-11-22  
; PRIOR APPLICATION NUMBER: JAPAN 2000-149106

/	PRIOR FILING DATE:	2000-05-19					
/	NUMBER OF SEQ ID NOS:	3					
/	SOFTWARE:	PatentIn Ver. 2.1					
/	SEQ ID NO 1						
/	LENGTH:	4310					
/	TYPE:	DNA					
/	ORGANISM:	Oryza sativa					
/	FEATURE:						
/	NAME/KEY:	CDS					
/	LOCATION:	(655)..(3828)					
/	US-10-447-135--1						
			Query Match	100.0%;	Score 3171;	DB 15; Length 4310;	
			Best Local Similarity	100.0%;	Pred. No. 0;		
			Matches 3171;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;	
Qy	1	ATCGAGATTGTTCCAGTAGATCAGZAGGGAGCTGCTGTGTTTGGGAGCAACTGTATGCTT	60				
Db	655	ATCGAGATTGTTCCAGTAGATCAGGAGGGAGCTGCTGTGTTTGGGAGCAACTGTATGCTT	714				
Qy	61	GCTCGTGTTGGAACTGGTGCTGTAGCGCCAGTCGTCGAGCTGCACGACGCCCTCGTCAG	120				
Db	715	GCTCGTGTTGGAACTGGTGCTGTAGCGCCAGTCGTCGAGCTGCACGACGCCCTCGTCAG	774				
Qy	121	GATCGACGCCCTGAAGCTGGTGTAGACGAAACCGGCAACAACCAATGCGAGCATTTCTTC	180				
Db	775	GATCGACGCCCTGAAGCTGGTGTAGACGAAACCGGCAACAACCAATGCGAGCATTTCTTC	834				
Qy	181	ATAAGAGGGTATCTCTCTCTCTCAGAGAAGATCCAAAATTCTGCTCTCTATCTCGG	240				
Db	835	ATAAGAGGGTATCTCTCTCTCTCAGAGAAGATCCAAAATTCTGCTCTCTATCTCGG	894				
Qy	241	ATTTTCCATGACCAGAAAAAATGTGTAGAACCAAAGCTAGTTCAAGCCCATTTCGTGA	300				
Db	895	ATTTTCCATGACCAGAAAAAATGTGTAGAACCAAAGCTAGTTCAAGCCCATTTCGTGA	954				
Qy	301	GCAAGTTTCGACGATGGGATTGCTCGAAGTGTCTTGATAGTTGAAACTTCAGATAT	360				
Db	955	GCAAGTTTCGACGATGGGATTGCTCGAAGTGTCTTGATAGTTGAAACTTCAGATAT	1014				
Qy	361	GGAAACAGCAACAAGAACTCTTCCCGCAAGACGAATGGCACAGTGAATGGTTCTCCATC	420				
Db	1015	GGAAACAGCAACAAGAACTCTTCCCGCAAGACGAATGGCACAGTGAATGGTTCTCCATC	1074				
Qy	421	ACATTTGTTCCGAGCACTTTTGTGCTGCTGTAGTGTTGGTTCCAAAAGTGCTCTTAGC	480				
Db	1075	ACATTTGTTCCGAGCACTTTTGTGCTGCTGTAGTGTTGGTTCCAAAAGTGCTCTTAGC	1134				
Qy	481	ACCAATCATCTCAAGGAGAGATGCTGTATAGATCAACTCTTCCAAGAGTGTGCAAGAA	540				
Db	1135	ACCAATCATCTCAAGGAGAGATGCTGTATAGATCAACTCTTCCAAGAGTGTGCAAGAA	1194				
Qy	541	GGCAATGACTCCAAATGCAATGCGCCTTCTGGCAAGATGGAGCTGCTGAGGCCAATACT	600				
Db	1195	GGCAATGACTCCAAATGCAATGCGCCTTCTGGCAAGATGGAGCTGCTGAGGCCAATACT	1254				
Qy	601	GATTCCACCATGAAAGATTGCAAGGCCAGCCCAAAATTATGATCTGCGACGAAATGTC	660				
Db	1255	GATTCCACCATGAAAGATTGCAAGGCCAGCCCAAAATTATGATCTGCGACGAAATGTC	1314				
Qy	661	TCTGAGCAACAACCTTCTGTGTATGTTGGGGCTTTTACCTGAACTTCCCGAGATTACATGG	720				
Db	1315	TCTGAGCAACAACCTTCTGTGTATGTTGGGGCTTTTACCTGAACTTCCCGAGATTACATGG	1374				
Qy	721	CACATGAGTAATAATGGTGCAGATCAACTCCATCCACTCCAAAATTTTCTGAAGTGTC	780				
Db	1375	CACATGAGTAATAATGGTGCAGATCAACTCCACTCCAAAATTTTCTGAAGTGTC	1434				
Qy	781	CTCAAAAGAAATCAAGATGAAAAATGAAAAATCTGAAGAGACTCTTGTGCTGAGCAGTGC	840				
Db	1435	CTCAAAAGAAATCAAGATGAAAAATGAAAAATCTGAAGAGACTCTTGTGCTGAGCAGTGC	1494				
Qy	841	AATTTGCCAAAGATVCTTAACCCAAATGTCTGGAAGAAACCTGTATCAGTTGCTGAGCAG	900				

1495 ATTTGACCAAGATCTTAACCAATGCTGGAAGGACGTGATCAGGTTCTTGACGAG 1554  
901 TCCAAATTTGACCAAGATCCGAAACAGATGCTGGGACAGAAATGTGACGAGATCTGCAAT 960  
1555 TCCAAATTTGACCAAGATCCGAAACAGATGCTGGGACAGAAATGTGACGAGATCTGCAAT 1614  
961 GAGCCATGTGAAGAGTGTCTTCTCAAAAGAAAGCTCCAAATCTTAAGAGGAAAGCGATTAAG 1020  
1615 GAGCCATGTGAAGAGTGTCTTCTCAAAAGAAAGCTCCAAATCTTAAGAGGAGGAGCGATTAAG 1674  
1021 AAGTTGATGAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1080  
1675 AAGTTGATGAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1734  
1081 GCAAGAGCTTTGTCGGAGGAGGACCAAAAGAGTGGGCTCTTATCAGAAATTAATAATGCT 1140  
1735 GCAAGAGCTTTGTCGGAGGAGGACCAAAAGAGTGGGCTCTTATCAGAAATTAATAATGCT 1794  
1141 AACCAAGTTGAGGATTTCTAGAAGTGAAGAAATTCATGCTGAAATTCGCGCTGATCCCTGT 1200  
1795 AACCAAGTTGAGGATTTCTAGAAGTGAAGAAATTCATGCTGAAATTCGCGCTGATCCCTGT 1854  
1201 GAGGATGATGAAGATACCATCCGCTCCGATGGAGTGAAGATGATGATGATGATGATGATGATG 1260  
1855 GAGGATGATGAAGATACCATCCGCTCCGATGGAGTGAAGATGATGATGATGATGATGATGATG 1914  
1261 AACCATACAGTGGGAGAGATGGGTTAAATCAAGTGAAGAAACAGACAAACGCAAAATAC 1320  
1915 AACCATACAGTGGGAGAGATGGGTTAAATCAAGTGAAGAAACAGACAAACGCAAAATAC 1974  
1321 TCTGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
1975 TCTGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2034  
1381 ACTGGAAGTGTGCATCACACAGTGTGCTCATCAGCTGGGAAATTCAGCAACCAAAAGAGT 1440  
2035 ACTGGAAGTGTGCATCACACAGTGTGCTCATCAGCTGGGAAATTCAGCAACCAAAAGAGT 2094  
1441 ACACCACTGCGAGTACTCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500  
2095 ACACCACTGCGAGTACTCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2154  
1501 AATATCATAAGACAGATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1560  
2155 AATATGATAGACAGATGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2214  
1561 TCAAGGGGAAACAGCGGTTTGATTAAGGGGAAACACATTCAGCTGCTAGTACCAAA 1620  
2215 TCAAGGGGAAACAGCGGTTTGATTAAGGGGAAACACATTCAGCTGCTAGTACCAAA 2274  
1621 TATGTTGTTGAAGACACAGGATGCTCAGAACATACATGATGATGATGATGATGATGATGATG 1680  
2275 TATGTTGTTGAAGACACAGGATGCTCAGAACATACATGATGATGATGATGATGATGATGATG 2334  
1681 TGCAGATGGAACCGAAACCTGTGTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1740  
2335 TGCAGATGGAACCGAAACCTGTGTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 2394  
1741 CATGATATCCAAATATGCTGACCTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1800  
2395 CATGATATCCAAATATGCTGACCTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 2454  
1801 CAAAACTTTGAAGTGAATCTCGTCAAAACAGACCATGATGATGATGATGATGATGATGATGATG 1860  
2455 CAAAACTTTGAAGTGAATCTCGTCAAAACAGACCATGATGATGATGATGATGATGATGATGATG 2514  
1861 GTTGAATGCTAGCTAAAAACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1920  
2515 GTTGAATGCTAGCTAAAAACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2574  
1921 GACATCAACCGTATTCAATCCAGACAACTGCTGATGATGATGATGATGATGATGATGATGATGATG 1980

2575 GACATCAACCGTATTCAATCCAGACAACTGCTGATGATGATGATGATGATGATGATGATGATG 2634  
1981 AAGGATCGTTGAGATTAATGATCAAGTGTGTTTGCACACTAATCCCAACAGAGTCCCTTG 2040  
2635 AAGGATCGTTGAGATTAATGATCAAGTGTGTTTGCACACTAATCCCAACAGAGTCCCTTG 2694  
2041 GCATCCCAAGATACACAGAGGAGTTACAGGCTCATTTGGCATTTGACACACACAGAGTCT 2100  
2695 GCATCCCAAGATACACAGAGGAGTTACAGGCTCATTTGGCATTTGACACACACAGAGTCT 2754  
2101 CCACATCTCAGAACTTTTCACTTCTACAGCAGCAGACACACATTTGCGATGCGAGAA 2160  
2755 CCACATCTCAGAACTTTTCACTTCTACAGCAGCAGACACACATTTGCGATGCGAGAA 2814  
2161 ATGTCCTACTATTTGTCGAAGCTCACCACTATTTTTCATCATCATGATGATGATGATGATGATG 2220  
2815 ATGTCCTACTATTTGTCGAAGCTCACCACTATTTTTCATCATCATGATGATGATGATGATGATG 2874  
2221 GAAGCACTCACTGAACTTTGGGCGGTAAAGGACCAAGAGCTAAAGTGGGACCAATTT 2280  
2875 GAAGCACTCACTGAACTTTGGGCGGTAAAGGACCAAGAGCTAAAGTGGGACCAATTT 2934  
2281 AAGGCCACTACAAGAAATTTCTCAGCAGCAGCAACATGCTGCTCAATTTAGACCTGGTATC 2340  
2935 AAGGCCACTACAAGAAATTTCTCAGCAGCAGCAACATGCTGCTCAATTTAGACCTGGTATC 2994  
2341 CAAGCAGTTGATTTGATCTTCTACTCATGTCATGGGATCTTCCAGCAATTTAGCATCTGCG 2400  
2995 CAAGCAGTTGATTTGATCTTCTACTCATGTCATGGGATCTTCCAGCAATTTAGCATCTGCG 3054  
2401 CAACCAAGTAAATTTGGGCACTGGACCGCTATGCTGAAAGAGCGGTAAACCAAGTCCATGCA 2460  
3055 CAACCAAGTAAATTTGGGCACTGGACCGCTATGCTGAAAGAGCGGTAAACCAAGTCCATGCA 3114  
2461 AGAAATTTTCCAGCACTATGACCAACCAATGGAAGCGAGTAAATTTGATGATGATGATGATGATG 2520  
3115 AGAAATTTTCCAGCACTATGACCAACCAATGGAAGCGAGTAAATTTGATGATGATGATGATGATG 3174  
2521 GCTGGAAGAGTGTGATCTTAAAGAAATCCATGCTGCGACGCTCTCTCAGAAATG 2580  
3175 GCTGGAAGAGTGTGATCTTAAAGAAATCCATGCTGCGACGCTCTCTCAGAAATG 3234  
2581 ATGATCCATCAACATTTAGCAAGCTTCCCAACTATGGAATCTTACAGGAAACAGATG 2640  
3235 ATGATCCATCAACATTTAGCAAGCTTCCCAACTATGGAATCTTACAGGAAACAGATG 3294  
2641 GAGTCTCACTTCAATTTCTCAGTATGCACTAATCAGTACAAAGGATCAACCAAGCA 2700  
3295 GAGTCTCACTTCAATTTCTCAGTATGCACTAATCAGTACAAAGGATCAACCAAGCA 3354  
2701 TCATATGCGAGTAACTGATGGAAGATTTCCATTTGACATTTGGAAGCTTTATCAGCGCAT 2760  
3355 TCATATGCGAGTAACTGATGGAAGATTTCCATTTGACATTTGGAAGCTTTATCAGCGCAT 3414  
2761 CAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2820  
3415 CAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3474  
2821 TTGCTGCAAGAGGAAATTTGCAAACTGGTGGAGAACTGTGGCAACCAATTTCTGTTATAAG 2880  
3475 TTGCTGCAAGAGGAAATTTGCAAACTGGTGGAGAACTGTGGCAACCAATTTCTGTTATAAG 3534  
2881 TTAAGAGTGTCAA CAGGAAATACATGCACTGATGAAACAGAGGAGGACATTTTGAAGCC 2940  
3535 TTAAGAGTGTCAA CAGGAAATACATGCACTGATGAAACAGAGGAGGACATTTTGAAGCC 3594  
2941 CTGAATTTCTGAAATTTTTCAGCAAAATGGAATGCAATTTGAGTTGGTCTTCTGTTAGCTCC 3000  
3595 CTGAATTTCTGAAATTTTTCAGCAAAATGGAATGCAATTTGAGTTGGTCTTCTGTTAGCTCC 3654  
3001 AGTCCAGATTTTATCAGCAGGAGGAAACAGCATAGCTCAATCTTGGACAGAGGAGGAGG 3060  
3655 AGTCCAGATTTTATCAGCAGGAGGAAACAGCATAGCTCAATCTTGGACAGAGGAGGAGG 3714



QY 3061 AAAATGTTTCATCCCTGGATCGTTTGTGAGACAGGATATCTGTATACTAACAGAC 3120  
Db |||||  
QY 3715 AAAATGTTTCATCCCTGGATCGTTTGTGAGACAGGATATCTGTATACTAACAGAC 3774  
Db |||||  
QY 3121 CCAGCTGATTTTACTACAATCAGTAACGATAACAGATATGATTACCGC 3171  
Db |||||  
QY 3775 CCAGCTGATTTTACTACAATCAGTAACGATAACAGATATGATTACCGC 3825  
Db |||||  
RESULT 3  
US-10-447-135-3  
; Publication 3, Application US/10447135  
; Publication No. US20030199684A1  
; GENERAL INFORMATION:  
; APPLICANT: Hirochika, Hirohiko  
; APPLICANT: Yamazaki, Munee  
; APPLICANT: Miyao, Akio  
; TITLE OF INVENTION: A novel gene involved in brassinosteroid responses  
; FILE REFERENCE: NAF-1 DIV  
; CURRENT APPLICATION NUMBER: US/10/447,135  
; CURRENT FILING DATE: 2003-05-27  
; PRIOR FILING DATE: 2003-05-27  
; PRIOR FILING DATE: 2000-11-22  
; PRIOR APPLICATION NUMBER: JAPAN 2000-149106  
; PRIOR FILING DATE: 2000-05-19  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 9455  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; US-10-447-135-3  
Query Match 81.1%; Score 2573.2; DB 15; Length 9455;  
Best Local Similarity 87.8%; Pred. No. 0;  
Matches 3003; Conservative 0; Mismatches 23; Indels 396; Gaps 2;  
QY 146 ACGAACCGGCACACACCAATCGGAGCAATTCCTCCATAGAGGGTATGTTCTCTTC 205  
Db 5447 ATGCATCTTTATATGATTAATCTTATTTTCTTGATCCAGAGGGTATGTTCTCTTC 5506  
QY 206 AGAAGAGGATCCAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 265  
Db 5507 AGAAGAGGATCCAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5566  
QY 266 ATGAACACAAAGCTAGTTCAGGCCCAATTTTCTGTAGCAAAAGTTTCCAGATGGATGCT 325  
Db 5567 ATGAACACAAAGCTAGTTCAGGCCCAATTTTCTGTAGCAAAAGTTTCCAGATGGATGCT 5626  
QY 326 CGAAGTCCTTGGATAGTTGAACACTTCAGATATGGAACAGCAGCAGCAAGTCTTCCCG 385  
Db 5627 CGAAGTCCTTGGATAGTTGAACACTTCAGATATGGAACAGCAGCAGCAAGTCTTCCCG 5686  
QY 386 CAAAGCAGAAATGGCACAAGTGTGTTGCTCCATCACAATTTGTTCCGAGCACTTTTGTGC 445  
Db 5687 CAAAGCAGAAATGGCACAAGTGTGTTGCTCCATCACAATTTGTTCCGAGCACTTTTGTGC 5746  
QY 446 CTGCTAGTGTGTTTCCCAAAAGTGTCTCTAGCAGCAATCATCTCAAGGGAAGATG 505  
Db 5747 CTGCTAGTGTGTTTCCCAAAAGTGTCTCTAGCAGCAATCATCTCAAGGGAAGATG 5806  
QY 506 CTGATAGTCAACTCTTCCAAAGAGTGTGCAAGAGGCAATGACTCCAAATGCAATGCCG 565  
Db 5807 CTGATAGTCAACTCTTCCAAAGAGTGTGCAAGAGGCAATGACTCCAAATGCAATGCCG 5866  
QY 566 CTTCGGCAGAAATGGAGCTGTGAGGCCAATCTGATTCACCAATGAA----- 614  
Db 5867 CTTCGGCAGAAATGGAGCTGTGAGGCCAATCTGATTCACCAATGAAAGTATGGTAG 5926  
QY 615 ----- 614  
Db 5927 ATGTAGAGCCTTCAAAATCCTAAGTAGGATTTTATTTAAGGTATAGATTAATTAATGT 5986

QY 615 -----AGATTTGCAAGGCCACGCCCAAAATTTATGATGTCGCGCAAAATGT 659  
Db 5987 TTGTGTGATTTTCTCAGATTTGCAAGGCCACGCCCAAAATTTATGATGTCGCGCAAAATGT 6046  
QY 660 CTCTGAGGACAAACACTTCTGTTTGTGTTTGGGGCTTTTACCTGAAAGTTCCTCAGATTAATG 719  
Db 6047 CTCTGAGGACAAACACTTCTGTTTGTGTTTGGGGCTTTTACCTGAAAGTTCCTCAGATTAATG 6106  
QY 720 GCACATAGAGTAATGTTGTCGAGATCAACCTCCATCCACTCCAAACTTTCTTGAAGTGT 779  
Db 6107 GCACATAGAGTAATGTTGTCGAGATCAACCTCCATCCACTCCAAACTTTCTTGAAGTGT 6166  
QY 780 CCTCAAAAGAAATGAAGATGAAATGGAATAAATCTGAGAGACTCTTGTGTTGTCGAGAGT 839  
Db 6167 CCTCAAAAGAAATGAAGATGAAATGGAATAAATCTGAGAGACTCTTGTGTTGTCGAGAGT 6226  
QY 840 CAAATTTGACAAAGATCTTAACCCCAATGTCCTGGAAGGAACGTGATCAGTTGCTGAGCA 899  
Db 6227 CAAATTTGACAAAGATCTTAACCCCAATGTCCTGGAAGGAACGTGATCAGTTGCTGAGCA 6286  
QY 900 GTGCAATTTGACAAAGATCCGAAACCACTGTCCTGGCAGAAATGTGAGCAGATCTGCAA 959  
Db 6287 GTGCAATTTGACAAAGATCCGAAACCACTGTCCTGGCAGAAATGTGAGCAGATCTGCAA 6346  
QY 960 TGAGCCATGTGAAGATGTTTCTCAAAAGAAAGTCTTCAATCTTAAGAGGAAGACGATAA 1019  
Db 6347 TGAGCCATGTGAAGATGTTTCTCAAAAGAAAGTCTTCAATCTTAAGAGGAAGACGATAA 6406  
QY 1020 GAAGTTGATGAAGAGCAGCAGCAGCAAGAAACGCTGCTGCCAGGCTGATGTTTCAGA 1079  
Db 6407 GAAGTTGATGAAGAGCAGCAGCAGCAAGAAACGCTGCTGCCAGGCTGATGTTTCAGA 6466  
QY 1080 TGCAAGCTTTTCTCGAGAAAGCCCAAAAGTGTGCGGTCTTCTATCAGAAATTTATAATGC 1139  
Db 6467 TGCAAGCTTTTCTCGAGAAAGCCCAAAAGTGTGCGGTCTTCTATCAGAAATTTATAATGC 6526  
QY 1140 TAAACAGGTTGAGGATCTTCAAGTGTGCGAAGTTCATGCTGGAATTCGCTGATCCCTG 1199  
Db 6527 TAAACAGGTTGAGGATCTTCAAGTGTGCGAAGTTCATGCTGGAATTCGCTGATCCCTG 6586  
QY 1200 TGAGATGATGAAGTACCATCCCGTCCGATGCAAGTAAAGTATGATTTCTGTTAG 1259  
Db 6587 TGAGATGATGAAGTACCATCCCGTCCGATGCAAGTAAAGTATGATTTCTGTTAG 6646  
QY 1260 CAAACATACAGTGGGGAAGATGGGTTAAATCAAGTAAAGAAACAAGAAACGCAATA 1319  
Db 6647 CAAACATACAGTGGGGAAGATGGGTTAAATCAAGTAAAGAAACAAGAAACGCAATA 6706  
QY 1320 CTCTGATGTTGTAGATGATGATCATCTTATGAACTGGCTGAATGGAATAAAGAAAG 1379  
Db 6707 CTCTGATGTTGTAGATGATGATCATCTTATGAACTGGCTGAATGGAATAAAGAAAG 6766  
QY 1380 AACTGGAAGTGTGCATCACAGTTGCTCATCCAGCTGGGAATTTGAGCAACAAAAGT 1439  
Db 6767 AACTGGAAGTGTGCATCACAGTTGCTCATCCAGCTGGGAATTTGAGCAACAAAAGT 6826  
QY 1440 GACACCCCTGCGAGTACTCAGCATGTATGATGAGATGATGATGATGATGATGATGATGAT 1499  
Db 6827 GACACCCCTGCGAGTACTCAGCATGTATGATGAGATGATGATGATGATGATGATGATGAT 6886  
QY 1500 AAATATGCAATAGACAGATGCTGTGTCAGCATGTATCAGAAATCTCCACACAGAGTGTCT 1559  
Db 6887 AAATATGCAATAGACAGATGCTGTGTCAGCATGTATCAGAAATCTCCACACAGAGTGTCT 6946  
QY 1560 ATCAAGGGGAAAACAGCGGGTTTGAATGAGGGAACACATTCGCTGTAGTACCAA 1619  
Db 6947 ATCAAGGGGAAAACAGCGGGTTTGAATGAGGGAACACATTCGCTGTAGTACCAA 7006  
QY 1620 ATATGTTGTGAAAGCAGCAGAAATGGTTCAGAAATACATATGATGATCAGCGCAGAGATCA 1679  
Db 7007 ATATGTTGTGAAAGCAGCAGAAATGGTTCAGAAATACATATGATGATCAGCGCAGAGATCA 7066



2720	ATGGAAGATTCCATTGTACATTCGAAGACTTATCACGSCATCA--GCTGCATGATCTGC	2776	
Db	105	ATGGAAGCCAACCGCTGAACTTGAAGACTTGTCTCGSGCTCATTCAGCAAGACTTGC	164
2777	ACAGACCTTTAGGCCACCATCTTAGAGTTGGTGTGCTTGGCTCTCTTCTGCAGAGAGAA	2836	
Db	165	GAAGGCTTTTACGNCCTCACCCCTGANTCGGTGTGCTCGGTTCATTGCTGCAGCAGGAGA	224
2837	TTGCAAACTGTCGGAGAACTGTGGCACACAATCTGGTTATAAGTTAGGAG	2897	
Db	225	TCGCAAACTGTCCTGGGAA--TGCGGCGCCGAGTCTGGGTANAGACTAGTG	274

```

RESULT 6
US-10-398-221-7/c
; Sequence 7, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIORITY APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 319630
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-7

```

Query Match	1.3%;	Score 42.2;	DB 16;	Length 319630;
Best Local Similarity	47.0%;	Pred. No. 32;		
Matches 131;	Conservative 0;	Mismatches 148;	Indels 0;	Gaps 0;
574	AAAGATGGAGCTGCTGGAGCCGCAATCTGATTCACCAATGAAGATTTGCAAGGCCAGCC	633		
289425	AACAAAAGAGCATGAATTTATGCCCAATATGCCAACAGAGAAGTTTCTGTGTGTGCG	289366		
634	CAAAATTATGATGGCGACGCAAGTCTCTGAGGACCAACCTTCTGTGTGATTTGGGGCT	693		
289365	GATPACTAAAGTGGAAAGGTTATGTTTCAAGCAGGAAACCACTTAGTTATGCHAGGAAT	289300		
694	TTACCTGAAGTTCGCCAGATTACATGGCCACATGAAGTAAATGGTGCAGATCAACCTCCA	753		
289305	ATTATTGATGACTTCAAAATCACTTCGCAAAAGGTGGTATTGTGTGGCGTTGGAAGCTGCA	289244		
754	TCCACTCCAAACCTTTCTGAGTGGTCTCTCAAAAGAAATGAAGATGAAAATGGAAGAACT	813		
289245	TCTGGCHAGAAATTTTAAAGATTTAATGCCACAGATGAGGTTCTCATTTATTAGGC	289118		
814	GAAGAGACTCTTGTGTGACGACGTGCAATTTGACCAA	852		
289185	GAAGTGGCTCTAGTTCCAGATCCCATCCCTATTTTCCCAA	289147		

```

RESULT 7
US101398-321-2058
; SOURCE: 2058, Application US/101398221
; Publication No. US3004018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua,
; genome and applications

```

FILE REFERENCE: 344 702 - US  
CURRENT APPLICATION NUMBER: US/10/398,221  
PRIOR FILING DATE: 2003-03-27  
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061  
PRIOR FILING DATE: 2001-10-04  
PRIOR APPLICATION NUMBER: FR 00/12 697  
PRIOR FILING DATE: 2000-10-04  
NUMBER OF SEQ ID NOS: 4025  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2058  
LENGTH: 3011208  
TYPE: DNA  
ORGANISM: Listeria innocua  
US-10-398-221-2058

Query Match 1.3%; Score 42.2; DB 16; Length 3011208;  
Best Local Similarity 47.0%; Pred. No. 1.6e+02;  
Matches 131; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 574 AGAATGGAGCTCTGAGGCGCAATCTGATTCACCAATGAAGATTGCAAGGGCCAGCC 633  
DB 1827998 AACAAAAAGGACATGAATTTATGGCCAAATATGCCAACAGAGAGTTCTTCTGTTGGCG 1828057  
QY 634 CAAATTTATGCTGCGCAGCAATGCTCTGAGGACACACTTCTGTTGATGTTGGGGCT 693  
DB 1828058 GATRAACTAAAGTGAAGTTATGTTCAAGCAGAACCACTTAGTTATGCGAGAAAT 1828117  
QY 694 TTACCTGAAGTCCCGCAGATTACATGCGCACATAGAAAGTAAATGGTGCAGATCAACTCCA 753  
DB 1828118 ATTATTGATGACTTCAAAATCACTTTCGAAAAGGTCGTATTGTTGGCGTTGAAAGTCCA 1828177  
QY 754 TCCACTCCAAAAGCTTCTGAGGCGGCTCTCAAAAGAAATGAAGATGAAATGAAAAAAT 813  
DB 1828178 TCTGGCGAGAAATTTTAAAGATTTAATCGCAAGATGAGGTTCTCATTTATTAGCG 1828237  
QY 814 GAAGACACTTTGTTCTGAGCAGTCCAAATTTGACCAA 852  
DB 1828238 GAAGTGGCTAGTTCAGATCCATCCCTATTTCCAA 1828276

## RESULT 8

US-10-282-122A-34801  
Sequence 34801, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23

## RESULT 9

US-10-188-186-59  
Sequence 59, Application US/10188186  
Publication No. US20040029789A1  
GENERAL INFORMATION:  
APPLICANT: Anderson et al.  
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
FILE REFERENCE: 21402-397C  
CURRENT APPLICATION NUMBER: US/10/188,186  
CURRENT FILING DATE: 2002-07-02  
PRIOR APPLICATION NUMBER: 60/303046  
PRIOR FILING DATE: 2001-07-05  
PRIOR APPLICATION NUMBER: 60/360814  
PRIOR FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: 60/303828  
PRIOR FILING DATE: 2001-09-07  
PRIOR APPLICATION NUMBER: 60/323380  
PRIOR FILING DATE: 2001-09-19  
PRIOR APPLICATION NUMBER: 60/361133  
PRIOR FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: 60/304016  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: 60/304502  
PRIOR FILING DATE: 2001-07-11  
PRIOR APPLICATION NUMBER: 60/305262  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: 60/373881  
PRIOR FILING DATE: 2002-04-19  
PRIOR APPLICATION NUMBER: 60/305673  
PRIOR FILING DATE: 2001-07-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 368  
SOFTWARE: Custom  
SEQ ID NO 59  
LENGTH: 5451

PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 34801  
LENGTH: 1260  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-10-282-122A-34801

Query Match 1.3%; Score 41; DB 13; Length 1260;  
Best Local Similarity 49.8%; Pred. No. 1.2;  
Matches 104; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 473 CTCCTAGCACCAATCATCTCAAGGGAAGAAATGCTGATGATCAACTCTTCCAAAGAGTG 532  
DB 1052 CTATAAGTGAAGAAAGCAATCGAAGCTAAACAGGAGCGCTGATTCAGTTCAATTATAG 1111  
QY 533 TGCAGAGGCAATCACTCCAAATGCAATGCGCTTCTGGCAAGATCGAGCTGCTGAGG 592  
DB 1112 AAGAGCAATTAATTGACATCATGATGATGATGATCTTCTCGAAAATGTAAGTAAAGTAG 1171  
QY 593 CCAATACTGATTCCCAATGAAGATTGCAAGGCGCAGCCCAAAATATGATGTTGGCAG 652  
DB 1172 TCATTACTGAACAAACAAATTAATGAAGAAATTGAGCCTGATTTATGATGATGAAGAA 1231  
QY 653 CAAATGCTCTGAGGACACACTTCTGTT 681  
DB 1232 ATCTAATTAATGAAGCAACAGACATCTGCT 1260

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (163)..(5146)
US-10-188-186-59

Query Match
  1.3%; Score 40.6; DB 13; Length 5451;
Best Local Similarity 48.9%; Pred. No. 4.7; Indels 0; Gaps 0;
Matches 109; Conservative 0; Mismatches 114;

Qy 1716 CTCGGCAAGGTTCTCCAGCTGAGCATGATATCCAAATTTATGTCGACCTTCATGAGCA 1775
Db 2235 CTCGGTAGAATCTTACTCAGAGAAAGCAAAAAGTTAGTCTGATATGAAAGAA 2294

Qy 1776 GAGTCTACCAAGAAAGAGCAAACTTGAAGTCACTCGTGAAACAGACCAT 1835
Db 2295 ACTCCAAGCCATGAACAAGAACTGCAGACATTCAGCGAGCTCAAAAAGACATGCAAG 2354

Qy 1836 GATAGATGATATCCCATGATATTTGTAAGTCTAGCTTAAACACAGCATGAGAGCA 1895
Db 2355 GTTGCTTAAATATCAGTCTCAGTATGAAAGCAATTTGAAGAAATTCAGCAGCATGTGAT 2414

Qy 1896 GCTTATGATGAGACTGATTTGTTCTGACATCAACCGTATTCAA 1938
Db 2415 GGAATGAAAAAACAAGGTTCCGCTTAATGAAACAAATGAAA 2457

RESULT 10
US-10-302-172-23
; Sequence 23, Application US/10302172
; Publication No. US20040053250A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20040053250A1el Arginine-rich Protein-like Nucleic Acids an
; FILE REFERENCE: 803-1CNCB
; CURRENT APPLICATION NUMBER: US/10/302,172
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/225,251
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: PCT US02/05095
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 950
; SOFTWARE: pt_FL_genes Version 2.0
; SEQ ID NO 23
; LENGTH: 5761
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (163)..(5148)
US-10-302-172-23

Query Match
  1.3%; Score 40.6; DB 13; Length 5761;
Best Local Similarity 48.9%; Pred. No. 4.9;
Matches 109; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 1716 CTCGGCAAGGTTCTCCAGCTGAGCATGATATCCAAATTTATGTCGACCTTCATGAGCA 1775
Db 2235 CTCGGTAGAATCTTACTCAGAGAAAGCAAAAAGTTAGTCTGATATGAAAGAA 2294

Qy 1776 GAGTCTACCAAGAAAGAGCAAACTTGAAGTCACTCGTGAAACAGACCAT 1835
Db 2295 ACTCCAAGCCATGAACAAGAACTGCAGACATTCAGCGAGCTCAAAAAGACATGCAAG 2354

Qy 1836 GATAGATGATATCCCATGATATTTGTAAGTCTAGCTTAAACACAGCATGAGAGCA 1895
Db 2355 GTTGCTTAAATATCAGTCTCAGTATGAAAGCAATTTGAAGAAATTCAGCAGCATGTGAT 2414
```

```
Qy 1896 GCTTATGATGAGACTGATTTGTTCTGACATCAACCGTATTCAA 1938
Db 2415 GGAATGAAAAAACAAGGTTCCGCTTAATGAAACAAATGAAA 2457

RESULT 11
US-09-918-995-22363
; Sequence 22363, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22363
; LENGTH: 412
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-918-995-22363

Query Match
  1.3%; Score 40; DB 10; Length 412;
Best Local Similarity 48.3%; Pred. No. 1.1;
Matches 112; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Qy 1247 ATATTCTCTTTAGCAACCATACAGTGGGAGAGATGGGTTAAATCAAGTAGACACA 1306
Db 168 ACATCCAGGAACCCAGGCTTAGAGAGGGAACATGGACCTTTGGAGGGAGCTACCTAC 227

Qy 1307 CAAAAGCGAAATACCTCTGATGTTGTAGATGATGGATCATCTATGTAACCTGGCTGAATG 1366
Db 228 CAGGTGCCAGATGCCCTCGAGTGGGACAGAGATCAAGGCTTGTGAGAGCTGGATG 287

Qy 1367 GAAAAAGAAAAAGAACTGGAAGTGTGATCAACAGCTGCTCATCCAGCTGGGAATTCGA 1426
Db 288 GAAAAATTTAAAGACAAGGCTGGTTACATGTCTGGGATGCTGCTGCTGGGGTGGG 347

Qy 1427 GCACACAAAAAGTGCACACCCACTGCGAGTACTCAGCATGATGAGATGA 1478
Db 348 ATAGCTGGAGCCTTGTTCATCTTGGAGCCCTCTACAGCATTAAGGTTATGA 399

RESULT 12
US-10-311-455-748/c
; Sequence 748, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIERENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 748
; LENGTH: 15122
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
```

RESULT 15  
US-09-745-763-114  
Sequence 114, Application US/09745763  
Patent No. US20020065394A1  
GENERAL INFORMATION:  
APPLICANT: Jacobson, Kenneth  
McCoy, John M.  
Lavallie, Edward R.  
Collins-Racie, Lisa A.  
Evans, Cheryl  
Merberg, David  
Treacy, Maurice  
Spaulding, Vikki  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
NUMBER OF SEQUENCES: 219  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/745,763  
FILING DATE: 18-Jun-2000  
CLASSIFICATION: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284

Search completed: May 17, 2004, 05:12:31  
Job time : 1332 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 16, 2004, 19:11:26 ; Search time 7740 Seconds  
(without alignments)  
12234.234 Million cell updates/sec

Title: US-09-721-114-1\_COPY\_655\_3825

Perfect score: 3171  
Sequence: 1 atggagattgttcagtaga.....acgagtatatggattaccgc 3171

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gsa1:\*  
29: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	355.6	11.2	1025	28	BZ795917 PUFHM45TD
2	351.2	11.1	888	28	CC355521 PUFHPL93TB
3	268.2	8.5	523	12	BJ246359 BJ246359
4	237.4	7.5	885	29	CG091207 PUJFK24TD

5	236	7.4	469	29	AG213697
6	229.6	7.2	593	13	BQ244810
7	205.8	6.5	999	28	BZ795914
8	205.2	6.5	556	13	BU974305
9	203	6.5	787	29	CG345663
10	178.6	5.6	469	9	AL818058
11	171.2	5.4	491	29	AG209720
12	168.8	5.3	641	28	BZ411511
13	167.2	5.3	849	28	BZ411504
14	152	4.8	766	29	CG603916
15	139.4	4.4	635	28	BH880296
16	129	4.1	822	29	CG385016
17	124.4	3.9	753	28	CC160109
18	124.4	3.9	995	29	CG451574
19	122.8	3.9	842	28	BZ705032
20	120.8	3.8	703	28	BZ776688
21	119.6	3.8	688	28	BZ776689
22	118.8	3.7	970	29	CG032005
23	114.2	3.6	653	28	BZ311518
24	114.2	3.6	870	29	CG091452
25	111.2	3.5	903	28	BZ704364
26	103.6	3.3	793	28	BZ423675
27	103.2	3.3	782	29	CC824592
28	98.6	3.1	701	28	BZ374744
29	91.2	2.9	590	28	CC160110
30	91	2.9	562	28	CC044201
31	91	2.9	792	28	BZ422300
32	91	2.9	809	28	BZ652757
33	83.8	2.6	905	29	CG365008
34	82.4	2.6	664	13	BQ242025
35	80.2	2.5	600	29	CC603904
36	78.2	2.5	183	12	BZ252288
37	78.2	2.5	389	29	CC991021
38	77.4	2.4	628	29	CC744988
39	77.4	2.4	734	28	BZ326387
40	76.6	2.4	632	13	BQ161601
41	76	2.4	587	28	BZ359680
42	75	2.4	321	9	AI691520
43	74	2.3	707	29	CG837058
44	72.6	2.3	562	28	BZ652753
45	67.2	2.1	373	29	CG831985

ALIGNMENTS

RESULT 1  
BZ795917/c  
LOCUS  
DEFINITION  
BZ795917  
VERSION  
BZ795917.1  
KEYWORDS  
GSS.  
SOURCE  
Zea mays  
ORGANISM  
Zea mays  
REFERENCE  
AUTHORS  
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennettren,J.  
TITLE  
JOURNAL  
COMMENT  
Maize Genomics Consortium  
Unpublished (2003)  
Other GSSs: PUFHM45TB  
Contact: Cathy Whitelaw  
TIGR Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TP

BZ795917 1025 bp DNA linear GSS 17-MAR-2003  
PUFHM45TD ZM.0.6.1.0 XB Zea mays genomic clone ZMMBTA320H17,  
genomic survey sequence.  
ACCESSION BZ795917  
VERSION BZ795917.1  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
REFERENCE  
AUTHORS  
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennettren,J.  
TITLE  
JOURNAL  
COMMENT  
Maize Genomics Consortium  
Unpublished (2003)  
Other GSSs: PUFHM45TB  
Contact: Cathy Whitelaw  
TIGR Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TP

BZ795917 1025 bp DNA linear GSS 17-MAR-2003  
PUFHM45TD ZM.0.6.1.0 XB Zea mays genomic clone ZMMBTA320H17,  
genomic survey sequence.  
ACCESSION BZ795917  
VERSION BZ795917.1  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
REFERENCE  
AUTHORS  
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennettren,J.  
TITLE  
JOURNAL  
COMMENT  
Maize Genomics Consortium  
Unpublished (2003)  
Other GSSs: PUFHM45TB  
Contact: Cathy Whitelaw  
TIGR Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TP

Class: sheared ends. Location/Qualifiers 3'..1025 /organism="Zea mays" /mol\_type="genomic DNA" /strain="B73" /db\_xref="taxon:4577" /clone\_lib="ZM0.6.1.0 KB" /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high COT selected genomic DNA library"

FEATURES source Query Match 11.1%; Score 355.6; DB 28; Length 1025; Best Local Similarity 67.0%; Pred. No. 5.5e-76; Matches 619; Conservative 0; Mismatches 254; Indels 51; Gaps 6;

ORIGIN

2099 CTCCACATCTCTGAGACTTTCAGTCTACCTACAGGAGCAGACACATTTGGCGATGGAG 2158

888 CTCTCTCTGATGACATTTTGGAGTGATCCACAGACACACAGACATTTCCACAGGGATG 829

2159 AAATGGTCACTATGCTGCAAGCTCACCACATTTTTCATCATCATGATGATGATGATG 2218

828 GAGGGGTACCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 775

2219 CTGAAGCACCACCTGACATTTGGGGCGGTAGGACGACGACGACGACGACGACGACG 2278

774 CTGAAGTGGCTGCTCGAGCTGAGGAAATACCGGGGAAAGAGTTAATGTGGGATCTT 715

2279 TTAAGGCCACTACAAGAAATTTCCAGCAGCAACATGCTGCTCAATTTAGACCTGGTA 2338

714 TCAGACAGCTTCAAGAAATTCACCAACATCGTCATATGTTTCAATTCGGAACAGC 655

2339 TCAAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2398

654 TTCAAGAAGTTTCAATCATCTCCAGTCTGCTTATGAGGCTTCTAGTAATCTATGAGCTC 595

2399 GCCAACCAAGTAATTTGGCCACTGCGCTATGCTGAAAGAGGGTTAACAGGTCATG 2458

594 ACCAGCAGTAATTTGGCTGTAGA-----TCAGC 565

2459 CAAGAAATTTTCAAGCACAATAGCAACCACTGGAAGCGATAGTTATGTATCGAGAA 2518

564 CGAGAAGTGTTCGAAGCACAACATGACTATGAGGCTGCTGATGATGATGATGATG 505

2519 ATGCTGACAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2578

504 TTCTTGACAGTCCAGGCTTACCCAAAGAGGCTATGCTGCAACACATCTTCTGAT 445

2579 TGATGATCCATCAACATTAGCAAGCTTCCCAACTATGGAATTTCTAGCAGGAACAGA 2638

444 TGATGACTCATCAACAGCTCCAGACTTCACAAACTATCAAGAGCTAACAGGCGCAGA 385

2639 TGGAGTCTCAACTTCA-----ATCTCAGTATGAC-----ATATCAGTACAAAGAT 2689

384 TGAACATCAACACAAACTCTGGATTCACATATACAGCATGATCAGTATATGGT 325

2690 CAACCAAGCAGCATCATATGCGAGTAAGCTGAATGGAAGATTCATTTGACATTCGAAGCT 2749

324 CACTGAGCAGCATCATATGGAAGACACATAACTGGAAGGTTCCACCGACACTGCAAGCT 265

2750 TATCAGGCAATGAG-----CTGCTGATGATGACAGAGCTTTTACGCCACATCTAGAGTTG 2806

264 TATCTGGGCTGAGTCCAGCAAACTGACAGAGGCTTTTACGCCCTTCTCTCGGGTGG 205

2807 GTGCTGTGGCTCTGCTGACAGAAATTCGAACTGCTGCGAGAACTGTGG---CA 2863

204 GCGTGTGTTGTTGCTGACAGAGATATCGCAACTGTTGATGAATGAACTGCGGCCGC 145

2864 CACAACTGTTTATAGTTAGGAGTGTCAACAGGAATAACATCGCATCATGATGAACAGAA 2923

144 CACAGTCTGATACAGCTGGGGCTTCTTAAAGGGAGCGCATCGCTTGTATGATGAACAGAG 85

2924 AGGAACATTTTGAAGCCCTGAATTCGGAATGTTTTCAGCAAAATGGAATGCAATTCAGT 2983

84 ATGGAACCTACGAGACCTTGAATTCAGGAGTGTTCACAGCAGGATGGAATGCCCTTCAAT 25

2984 TGGGTTCTGTTAGCTCCAGTGCAG 3007

24 TGGGTTCTGTTAGCTCTCTGCG 1

RESULT 2

CC355521 888 bp DNA linear GSS 16-MAY-2003

LOCUS PUHPL93TB ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMMBTAS04018,

DEFINITION genomic survey sequence.

ACCESSION CC355521

VERSION CC355521.1 GI:30824921

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE 1 (bases 1 to 888)

AUTHORS White, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J., and Benner, J.

TITLE Maize Genomics Consortium

JOURNAL Unpublished (2003)

COMMENT Other GSSs: PUHPL93TD

Contact: Cathy Whitelaw

TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers 1..888 /organism="Zea mays" /mol\_type="genomic DNA" /strain="B73" /db\_xref="taxon:4577" /clone\_lib="ZMMBTAS04018" /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high COT selected genomic DNA library"

ORIGIN

Query Match 11.1%; Score 351.2; DB 28; Length 888; Best Local Similarity 67.9%; Pred. NO. 6.3e-75; Matches 602; Conservative 0; Mismatches 233; Indels 51; Gaps 6;

2298 TTCTCCAGCAGCAGCATGCTGCTCAATTTAGACCTGATCCAGCAGTGTGATTCGAC 2357

1 TTCACCAACATCGTCATATGTTTCAATTCGGAACAGGCTTCAGAGTTTCAATCATC 60

2358 TTCTACTCATGTCATGGGATCTCCAGCAATTTATGCAATTCGCAACCAAGTAATTCGCC 2417

61 TCCAGTCACTGTTATGAGCTTCTAGTAATCATGAGCTCACCAGCCAGTAATTTGTGGC 120

2418 ACTGACCGCTATGCTGGAAGCGGTTAACAGGTCATGCAAGAAATTTTCCAGCAC 2477

121 TATAGT-----CAGCCGAGTGTTCAGGAC 150

2478 AATAGCAACCTGGAAGCGAGTAAGTTATGTATGTCGGAAGAAATGTCGACAGTATGCTT 2537

151 AACATCGACTATGAGGCTGTTAGTTTGTATGATCGAAGGATTTCTGACAGTCAAGGCT 210

2538 GTATCTTAAGATTCATGCTCGACGCTTTCTGAGATGATGATGATGATGATGATGATG 2597

211 CTACCCAAAGAGGCTATGCTGCTGCAACACATCTTCTGATGTTGAGTCTCATCAACAGC 270

QY 2598 AGAAGCTTCCCAACTATGNACTTCTAGCAGGAACAGATGGATCTCAACTTCATA- 2656  
Db 271 TCCAGACTTCCAAACTATCAAGAGCTAACAGGCGCCAGATGGAAATCAACACAAC 330  
QY 2657 -----AFTCTCAGTATGCA---CATATCAGTACAAAGGATCAACAGCATCATATGG 2708  
Db 331 TCTGGATTCAATATACAGAGATGATCAGTATATGGTCTACTGAGGACATCATATGG 390  
QY 2709 CAGTAACTGAAAGGATTCATTCAGATTCGAGATTCATTCAGGCTCAG---CT 2765  
Db 391 AAGACATCAATGGAAGGTTCCCGACATCGCAAGCTTATCTCGGGCTCAGGTCCA 450  
QY 2766 GCATGATCTGCACAGACCTTTAGGCCCACTATCTTAGAGTTGGTGTGCTTGGCTCCTTGGCT 2825  
Db 451 GCAAACTTCACAGAGCTTTAGGCCCTCATCTCGGGTGGCGTCTTGGTGGTGGCT 510  
QY 2826 GCAGAGGAATTCGAACTGGTGGGAGAACTGTGG---CACATATCTGGTTATAGTT 2882  
Db 511 GCAGCAGGATATCGCAAACTGGTATGAAACTGCGGGCCGCCACAGCTCTGGATACAGGCT 570  
QY 2883 AGGAGTCTCAACAGGATATCATCGCATCAGATGCAACAGAAAGCAATTTTGAAGCCCT 2942  
Db 571 GGGCTTCTTAAGGGAGGCAATCGCTTGATATGACAGACATGGAATACAGACCTT 630  
QY 2943 GAATCTGGGAATGTTTTCAGCAAAATGGAATGCAATGCAATGGTGGTCTCTTAGCT--- 2998  
Db 631 GAATCAGGAGTGTTCACAGCAGGATGGAATGCCCTTCAATGGGTTCTCTTAGCTCTCT 690  
QY 2999 --CCAGTGCAGATTTTATACAGCGAGACAGATAGCTCAATCTTGGACAGAGSCAA 3056  
Db 691 TGGCGATCCAGATACCGTGTCCAGGATGATGTAAGGTACGCTTGTACAGTGGCAA 750  
QY 3057 GGGTAAATGTTTCATCCCTTGGATCGGTTTGTGAGCAGGATATCTGTATATACTAACAA 3116  
Db 751 TGGCAGAACGGTTCAACCGTTGGATAGCTTGTGCGAAGGATATCTGTGACTACAG 810  
QY 3117 GAACCGAGCTGATTTACTACATCAGTACGATACAGATATATG 3162  
Db 811 AAACCCAGCTGATTTTACTGTATTAAGTACATGAAATGAGTATATG 856

RESULT 3  
LOCUS BJ246359 523 bp mRNA linear EST 05-APR-2002  
DEFINITION BJ246359 Y. Ogiwara unpublished cDNA library, Wh\_f Triticum aestivum cDNA clone whf22j14 5', mRNA sequence.  
ACCESSION BJ246359  
VERSION BJ246359.1 GI:20058037  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.  
AUTHORS Ogiwara, Y. and Murai, K.  
TITLES Expressed genes in Triticum aestivum  
JOURNAL Unpublished (2002)  
COMMENT Contact: Tadasu Shin-i  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers  
1. .523  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultiVar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="whf22j14"  
/tissue\_type="spike at flowering date"

ORIGIN  
Query Match 8.5%; Score 268.2; DB 12; Length 523;  
Best Local Similarity 72.9%; Pred. No. 1.3e-54;  
Matches 345; Conservative 0; Mismatches 128; Indels 0; Gaps 0;  
QY 2121 GTCTACTCAGGACAGACAGACATTTGGGATGGAAGAAATGGTCACTATTTCTGCTCAAG 2180  
Db 50 GTCTACCCAGGACACAGGACATTTGGGCGATGGGAGAAATAGTACATATTTCTGCAAC 109  
QY 2181 CTCACCACTATTTTTCATCATCATGATGATGATATTTGCTGAAGCAACCACTGAAATTTG 2240  
Db 110 CTCACCACTATTTATCACAGACAGAGGACGAGCTCTTCTGAAGCAACACCTGAGCGCTG 169  
QY 2241 GGGCGTGAAGAGCGCAACAGCACTTAACGTGGGAGCAATTTTAGGCCACCTACAGAATTC 2300  
Db 170 GAGCCATATTTGGGAGCAAGAGTCGATGTGGGAACCTTTCAAGGGCACTTTCCAGGGATTT 229  
QY 2301 TCCAGCAGCAACATGTGGTGGCTCAATTTTAGACTCTGTATCTCCAGCAGTTGACTTGACTTC 2360  
Db 230 ATCCACCAACAGAGTTGTGCTCAGTTTCAGACTAGCATTTGACACGGTTGATTTAACTTA 289  
QY 2361 TACTCATGTCAGGATCTTCCAGCAATTTATGATCTGCCCAACCAAGTAAATTTGCGCACT 2420  
Db 290 TACTCATGTCAGGAGGAGCTTAATCGTTTATTTATCCACTCGCCAGCCAGTAATTTGCACT 349  
QY 2421 GGACCGCTATGCTGAAAGAGCGGTTAAACCAAGTCCATGCAAGAAATTTTCCAGGCAAT 2480  
Db 350 TGACCACTATACAAATAGAGCAGTTAACTCAGTCCAGGCAAGAAATTTTCCAGTTCAGT 409  
QY 2481 AGCAACCAATGGAAGGAGTAAGTTATGATCGGAGAAATGCTGGCAAGTACTCTTGTGA 2540  
Db 410 GTCAACCAATGGAAGCTGGTAACTCTGTGTGATGGAAGAAATTTGGACATTCAGGTTTTTA 469  
QY 2541 TCTTAAGATTCATCCCTCGGAGCAGCTTTCTTGAGATGATGATGATCCCA 2593  
Db 470 TCCAGAGAAACCATGCTGCTGCTACTCGCCCTGAGATTTGACCGAGTCCCA 522

RESULT 4  
LOCUS CG091207 885 bp DNA linear GSS 20-AUG-2003  
DEFINITION PUJFK247D ZM.0.6.1.0 KB Zea mays genomic clone ZMBF0656C23, genomic survey sequence.  
ACCESSION CG091207  
VERSION CG091207.1 GI:33973501  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Panicoideae; Andropogoneae; Zea.  
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Bennetzen, J.  
TITLES Maize Genomics Consortium  
JOURNAL Unpublished (2003)  
COMMENT Other GSSs: PUJFK247B  
Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: sheared ends.  
Location/Qualifiers  
1. .885  
/organism="Zea mays"  
/mol\_type="genomic DNA"

1  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 source  
 1. 469  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /strain="ND6047"  
 /cultiivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="T10466T"  
 /clone\_lib="PCR product directly amplified from rice genomic DNA"  
 /note="The 3' end of retrotransposon Tos17 was found immediately upstream of this sequence."  
 1  
 misc\_feature  
 /note="This insertion point is base 1732 in AF326768."  
 ORIGIN  
 Query Match 7.4%; Score 236; DB 29; Length 469;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-47;  
 Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1492 CTTGACACAAATATCATGAAGACAGATGCTGTCTCAGCATGTATCAGAAATCTCAGAAATCTCCACACAG 1551  
 Db 1 CTTGACACAAATATGCATGAAGACAGATGCTGTCTCAGCATGTATCAGAAATCTCCACACAG 60  
 QY 1552 AGGTGCTCATCAAGGGGAAACACCGGGTTGAGTAGGGGAAACACATTCAGCTGCT 1611  
 Db 61 AGGTGCTCATCAAGGGGAAACACCGGGTTGAGTAGGGGAAACACATTCAGCTGCT 120  
 QY 1612 AGTACCAATATGTTGGTGAAGACACCAAGAAATGTCAGAACATACATGTTACTCAGCGCA 1671  
 Db 121 AGTACCAATATGTTGGTGAAGACACCAAGAAATGTCAGAACATACATGTTACTCAGCGCA 180  
 QY 1672 GAAGATCAATGTCAGATGAAGAACCGAAAACCTCTGTTCTGAGTCATCGCAAGGT 1727  
 Db 181 GAAGATCAATGTCAGATGAAGAACCGAAAACCTCTGTTCTGAGTCATCGCAAGGT 236  
 RESULT 6  
 BQ244810  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENTS

Contact: Dr. Sylvie Cloutier  
Cereal Research Centre, Agriculture and Agri-food Canada  
195 Dafoe Rd. Winnipeg, MB, Canada R3T 2M9  
Tel: (204) 983-2340  
Fax: (204) 983-4604  
Email: scloutier@agr.gc.ca

Was cloned directionally, not all sequences generated with reverse primer were from the 5' end (same with forward primer and 3' end).  
Average insert size is >1.4 kb  
Plate: 035 row: B column: 12  
Seq primer: M13 Reverse.

```

FEATURES
source
Location/Qualifiers
1..593
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TaG15035B12R"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"
/lab_host="E. coli DH10B"
/clone_lib="TaG15"
/note="vector: pCMV-SPORT6.0 (Invitrogen Technologies); Site 1: NotI; Site 2: MluI; mRNA obtained from wheat seeds of cultivar Glenlea 15 days post-anthesis"

```

ORIGIN

Of cultivar Grenada 13 days post-anthesis

Query Match 7.2%; Score 229.6; DB 13; Length 593;  
Best Local Similarity 74.0%; Pred. No. 3.9e-45;  
Matches 305; Conservative 0; Mismatches 104; Indels 3; Gaps 1;

Qy	2297	ATTCTCGAGCGAACATGTGGTGCTCTGAAATTTAGACCTGGTATCCAAAGCAGTGTGACTTGA	2356
Db	2	ATTTATCAAGCAACAAAGAGTTGTCTCAGTTCAGACCTAGCATTGACACGTTGAAITAA	61
Qy	2357	CTTCTACTCATGTCTATGGGATCTTCCAGCAATATTGCACTTCGCCAACCAAGTAAATTCGCG	2416
Db	62	CTTATACTGATGTGGCGGAGCTAATCGTTATTATCCACTTCGCCACCGCAGTAATTTTCGA	121
Qy	2417	CACITGGACCGGTATGCTTGAAGAGCGGTAAACCAAGTCCATGCAAGAAATTTTCCCAAGCA	2476
Db	122	CACITGACCACTATACAAATAGACGAGTTAACTCAGTCCAGGCAAGAAGTTTCCCAAGTT	181
Qy	2477	CAATAGCAACCATGGAAGCGAGTAAGTTATGTGATCGGAGAAATGCTGGGACAGTAGTCT	2536
Db	182	CAGTGTCAACCATGGAAGCTGGTAATCTGTGTGATGGAAGAAATGTTTGGACATTTCAGGTT	241
Qy	2537	TGTTATCCTTAAGATCCATGCTCGGAGCGATCTTCTGAGAAATGATGATCCATCAACAT	2596
Db	242	TTTATCAAGAGAAACCATGCTGCTACTCAGCCCTTGAGATGACCGAGTCACCAATGT	301
Qy	2597	TAGCAAGCTTCCCCAACTATATGGAATTTCTAGCAGGAACCAAGTGGAGTCTCAACTTCATA	2656
Db	302	TAGCCAGTTT---CAACTATGAAGGTCTAGCAGGAACCAAGTGAATTTCAACTTCGGA	358
Qy	2657	ATTCTCAGTATGCACATATACATGATCAAAAGGATCAACCAAGCACATCATATGG	2708
Db	359	ATTCACACTATGACAGGATCAGTCCATTCGGATCAGCCNGCACACCGTATGG	410

RESULT 7	BZ795914	999 bp	DNA	linear	GSS 17-MAR-2003
LOCUS	BZ795914				
DEFINITION	PUFM45TB ZM_0.6_1.0_KB Zea mays genomic clone ZMWBta320H17, genomic survey sequence.				
ACCESSION	BZ795914				
VERSION	BZ795914				
KEYWORDS	GSS.				
SOURCE	Zea mays				
ORGANISM	Zea mays				
	Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta; Spermatophyta, Magnoliophyta, Liliopsida, Poales, Poaceae, PACCAD clade, Panicoideae, Andropogoneae, Zea.				

```

REFERENCE
AUTHORS
    1 (bases 1 to 999)
    Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
    Resnick,A., Fraser,C.N., Yuan,Y., San Miguel,P., Ma,J. and
    Bennetzen,J.
TITLE
    Maize Genomics Consortium
JOURNAL
    Unpublished (2003)
COMMENT
    Other GSs: PUFHM45TD
    Contact: Cathy Whitelaw
TIGR
    9712 Medical Center Drive, Rockville, MD 20850, USA
    Tel: 301-838-5843
    Fax: 301-838-0208
    Email: whitelaw@tigr.org
    Seq primer: TR
    Class: sheared ends.
FEATURES
    source
        1..999
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="873"
            /db_xref="taxon:4577"
            /clone="ZMMBTrA320H17"
            /clone_lib="ZM_0.6_1.0 KB"
            /notes="vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
            Cot selected genomic DNA library"

```

ORIGIN	Query Match	6.5%; Score 205.8; DB 28; Length 999;
	Best local Similarity 63.8%; Pred. No. 3.3e-39;	
	Matches 400; Conservative 0; Mismatches 182; Indels 45; Gaps	
QY	2090 CACAGAGTCTCCACATCCTCAGAACCTTTCAGTCTACTCAGGAACAGCAGACACATTTGC 2149	
Db		
QY	409 CAAATTTACTGTCCTTGATGACATTTTGAGTGTTACCCAGAACACACAGACAAATTC 458	
Db		
QY	2150 GGAATGAAGAAATGGTCACTATTTCCTGCAAGCTTCACCACCTATTTTTCACATCATGATGATC 2209	
Db		
QY	469 ACAGGATGAGGGTCAACCATTGCATGTACCTCCTATGTTTTCATC 522	
Db		
QY	2210 AGTATATTGCTGAAGACCAACTGACATTTGGGGCGTGAAGACGCAAGAACTAACGT 2269	
Db		
QY	523 AACATATTGCTGAAGTGCCTGCTCGGAGCTGGAGGAATACCGGGAAAGAGTTAATGT 582	
Db		
QY	2270 GGGAGCAATTTAAGGCACCTACAGAAGAAATCTCCAGCAGCAACATGTGGTGTCTCAATTTA 2329	
Db		
QY	583 GGGATCTTTCAAGACAGCTTCAAGAAATTCACCAATCGTCATATGGTTTCAATTCG 642	
Db		
QY	2330 GACTGGTATCCAAAGCAGTTGACTTGCATTTCTACTCATGTTCATGGGATCTTCAGCAAT 2389	
Db		
QY	643 GAAACAGGCTTCAAGAGTTTCAATCTCTCCAGTCATGCTTATGGAGCTTCTAGTAAC 702	
Db		
QY	2390 ATGCATCTGCCAAACGATGTAATTTGGCGCACTGACCGCTATGCTGAAGAGCGGTTAAC 2449	
Db		
QY	703 ATGCAGCTCACACGCGAGTAATTTGGCTGTAGA 736	
Db		
QY	2450 AGGTCCATGCAAGAAATTTTCAAGCACAATAGCAACCATGGAAGCGAGTAAGTTATGTG 2509	
Db		
QY	737 -----TCAGCCGAGAGTGTTCCAGACCAACATCGACTATGGAGGCTGTGTAGTTGTATG 792	
Db		
QY	2510 ATCGAGAAATGCTGGCAAGTAGTCTTGATCTCTAAAGATCCATGCTCTGGCAGCGGATC 2569	
Db		
QY	793 ATCGAGGATTCCTGGCAGCTCAGGGCTCTACCCAAAAGAGCCCTATGCTCTGCAACACATC 852	
Db		
QY	2570 TTCTGAGATGATGGATCCATCAACATTAGCAGCTTCCCAACTATGGAACCTTCTAGCA 2629	
Db		
QY	853 TTCTGAGATGATGGACTCATCAACAGCTCCAGACTTCACAAATCTATCAAGAGCTTACA 912	
Db		
QY	2630 GGAACAGATGGAGTCTCAACTTCATA-----ATTCTCAGTAGCA---CATAACTCAGT 2680	
Db		
QY	913 GGGCCAGATGGAAACATCAACCAAACTCTGGATTCACAATATACACAGCATGATCAGT 972	
Db		
QY	2681 ACAAGATTCAAACGACATCATATG 2707	
Db		

[illegible]

2575 AGAATGATGATCATCAACATTTAGCAAGCTTCCCACTATGGAATCTCTAGCAGGAAC 2634  
 Db |||||  
 259 AGGCTGATGAGCTCATCAACAGCTCCAGGCTTCCAAACTTACCAAGAGCTAATAGGCG 318  
 Qy |||||  
 2635 CAGATGAGTCTCAACTTCTCATATTTCTCAGTATGACATCAATCAGTACAAAGATCAACC 2694  
 Db |||||  
 319 CAGATGAGCTTCAACAGCAAGCTCTGGCTCCAGTACACAGCATGAGCATTT----- 373  
 Qy |||||  
 2695 AGCACATCATATGCGAGTAACTGAAATGGAAGATTTCCATTTGACATTCGAAGCTTTATCA 2754  
 Db |||||  
 374 -ACAATGCGTCAACCGAGCAGCTCATATGGAAGCAACCGCTGACACTGGAAGACTTTGCT 432  
 Qy |||||  
 2755 CGGCATCA---GCTGATGATCTGCACAGACCTTTACGCCACATCTTAGAGTTGGTGG 2811  
 Db |||||  
 433 CGGGTCATTTCCAGCAAGACTTGGCAGGCTTTACGGCTTACCCCTGCTGTGTGGTGG 492  
 Qy |||||  
 2812 CTTGGCTCTTGTGTCAGAGGAAATTCGAACCTGTCGGAGAACTGTGGCACAATCT 2871  
 Db |||||  
 493 CTCGGTTTCATTTGTCGAGCAGGAGATCGCAACTGCTCTGGAACTGCGGGCGGAGTCT 552  
 Qy |||||  
 2872 GGTATAAGTTAGAGTGTCAACAGCAATAAATCCATCCATCAGATGAACAGGAACAT 2931  
 Db |||||  
 553 GGTACAGACTAGTGTATGTAAGGACAGGCCCTGTATGTCACAGAGCAGGAAC 612  
 Qy |||||  
 2932 TTGAAGCCCTGAATTTCTGGAATGTTTTCAGCAAAATGGAATGCCATTTGAGTTGGTTCT 2991  
 Db |||||  
 613 TACGAGACCTTGA-----GCTCGCAGGATGCAACACCCCTGCAGTTGGTTCT 660  
 Qy |||||  
 2992 GTTAGCT-----CCAGTGCAGATTTTATATCAGCAGCAAGACAGTACTCAATCTTGG 3045  
 Db |||||  
 661 GTTAGCTCTGTGCTCCAACTGAGATACCGGTTTCCATGATGATGATGATGATGATGATG 720  
 Qy |||||  
 3046 ACCAGAGCAAGGTTAAATGTTTCATCCCTTGGATCGGTTTGTGAGACAGGATATCTGT 3105  
 Db |||||  
 721 ACCAGTGCATATGGGAGGACCATTCATCCGTTGGATAGCTCGTGAGGAAGGATATCTGT 780  
 Qy |||||  
 3106 ATAACTA 3112  
 Db |||||  
 781 GTGACTA 787

RESULT 10  
 AL818058  
 LOCUS  
 DEFINITION  
 AL818058 j:223 Triticum aestivum cDNA clone E02\_j223\_plate\_11, mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Triticum aestivum (bread wheat)  
 Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Triticum.  
 REFERENCE  
 1 (bases 1 to 469)  
 AUTHORS  
 Wilson, I., Bewick, R., Shepherd, S., Barker, G., Parker, J., Owen, P.,  
 Edwards, D., Coghill, J., Holdsworth, M., Lenton, J., Shewry, P. and  
 Edwards, K.  
 TITLE  
 JOURNAL  
 COMMENT  
 A BSRRC-funded wheat EST resource for the academic community  
 Unpublished (2002)  
 Contact: Barker G  
 Institute of Arable Crop Research  
 Long Ashton, Bristol BS41 9AF United Kingdom.  
 FEATURES  
 source  
 1. .469  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="mercia"  
 /db\_xref="taxon:4565"  
 /clone="E02\_j223\_plate\_11"  
 /tissue\_type="embryo"  
 /dev\_stage="21 days post anthesis"  
 /clone\_lib="j:223"

ORIGIN  
 Query Match 5.6%; Score 178.6; DB 9; Length 469;  
 Best Local Similarity 65.2%; Pred. No. 1.2e-32;  
 Matches 306; Conservative 0; Mismatches 139; Indels 24; Gaps 2;  
 Qy 1960 GATTGTGTAATAGTAGTCTGCCAAGCATGTTTCAGATATGATGATCAAGTGTGTTGACACT 2019  
 Db 1 GATTGTGTAATAGTAGTCTGCCAAGCATGTTTCAGATATGATGATCAAGTGTGTTGACACC 60  
 Qy 2020 AATTCCTCCACAGAA---GTCTCTTGGCATCCCAAGTACACAGAAAGGTTTACAGGTTCT 2076  
 Db 61 ACTTCACAGCAGAAAGCGCTTTGGAACACAGATAGTTACCAAGATGCTTGCAGAAATGCT 120  
 Qy 2077 TTGGCATTTGACACACAAAG-----AGTCTCCACATCTTCAGAAC 2115  
 Db 121 GTATCAGCAGCAGCAGACACAGCTAGTATAGCATGCTCTCTCAAAATTTACAGACTCTTGGTCT 180  
 Qy 2116 TTTCAGTCTACTCAGAAACAGCAGACACATTTGCCGATGGAAGAAATGTTCACTATTGCT 2175  
 Db 181 GTAATGCTTACCAGGAACACAGACATTCGGSCATGGGAGATTAATGTTCACTATTGCT 240  
 Qy 2176 GCAAGCTCAGCAGCTATTTTTCATCATGATGATGATGATGATGATGATGATGATGATGATG 2235  
 Db 241 GCAAGCTCAGCAGCTATTTATCACCAGCAGGATCAGCTCTTGGCGAAGCAACCTGAG 300  
 Qy 2236 CATTTGGGCGCTGAAGAGCGCAAGAGCTAACGTGGGAGCAATTTAAGCCCACTACAAAGA 2295  
 Db 301 CCTGGAGCCATATTTGGAGCAAGAGTCTGATGGGAACCTTTCAAGGCACTTCCGAG 360  
 Qy 2296 AATTCCTCAGCAGCAACATGTTGGTCTCAATTTAGACCTGTTATCCAAAGCAGTTGACTTG 2355  
 Db 361 GATTATATCAACCAACAAAGGTTGCTCAGTTTCAGACCTAGCATTTGACACGTTTGATTTA 420  
 Qy 2356 ACTTCTACTCATGTCATGGGATCTTCCAGCAATATGCTATGCTATGCTATGCTATGCTATG 440  
 Db 421 ACTTATCTGATATGGGAGGACTAATGTTATTTATCCACTTCGCCAGC 469

RESULT 11  
 AG209720  
 LOCUS  
 DEFINITION  
 Oriza sativa (japonica cultivar-group) DNA, clone.NE8014.0.704\_1A,  
 3' flanking sequence of Tos17 insertion in rice strain NE8014,  
 genomic survey sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Oriza sativa (japonica cultivar-group)  
 Oriza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oriza.  
 REFERENCE  
 1  
 AUTHORS  
 Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K.,  
 Shinzuka, Y., Onosato, K. and Hirochika, H.  
 TITLE  
 Target Site Specificity of the Tos17 Retrotransposon Shows a  
 Preference for Insertion within Genes and against Insertion in  
 Retrotransposon-Rich Regions of the Genome  
 Plant Cell 15 (8), 1771-1780 (2003)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 2 (bases 1 to 491)  
 REFERENCE  
 12897251  
 AUTHORS  
 Miyao, A., Onosato, K. and Hirochika, H.  
 TITLE  
 Direct Submission  
 JOURNAL  
 FEATURES  
 source  
 1. .491  
 /organism="Oriza sativa (japonica cultivar-group)"



```
/mol_type="genomic DNA"
/strain="NE8014"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone_lib="pCR product directly amplified from rice genomic DNA"
/note="The 3' end of retrotransposon Tos17 was found immediately upstream of this sequence."
misc_feature
1
/note="This insertion point is base 139 in AF326768."
ORIGIN
Query Match
Best Local Similarity 5.4%; Score 171.2; DB 29; Length 491;
Matches 190; Conservative 0; Mismatches 15; Indels 1; Gaps 1;
QY 1 ATGAGATGTTGACGTA-GATCAGGAGGAGCTCGTGTGTGTGGACGCACTGTATGCT 59
DB 103 ATGAGATGTTGACGTAATGATCAGGAGGAGCTCGTGTGTGTGGACGCACTGTATGCT 162
QY 60 TGTCTGTGTGAGACTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 119
DB 163 TGTCTGTGTGAGAACTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 222
QY 120 GATGAGCGGCTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 179
DB 223 GATGAGCGGCTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 282
QY 180 CATAGAGGATGTTGCTCTTCTTC 205
DB 283 CATAGAGTATCATTTTCTGTATTTC 308
RESULT 12
BZ411511 641 bp DNA linear GSS 04-DEC-2002
LOCUS OGACC43TM ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0027112,
DEFINITION genomic survey sequence.
ACCESSION BZ411511.1 GI:26045013
VERSION BZ411511.1
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 641)
White, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGACC43TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1..641
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0027112"
/clone_lib="ZM 0.7 1.5 KB"
/note="Vector: pBCSK-; Site 1: HindII; 0.7-1.5 kb
methylation filtered genomic DNA library"
ORIGIN
/mol_type="genomic DNA"
/strain="NE8014"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone_lib="pCR product directly amplified from rice genomic DNA"
/note="The 3' end of retrotransposon Tos17 was found immediately upstream of this sequence."
misc_feature
1
/note="This insertion point is base 139 in AF326768."
ORIGIN
Query Match
Best Local Similarity 5.3%; Score 168.8; DB 28; Length 641;
Matches 299; Conservative 0; Mismatches 132; Indels 21; Gaps 3;
QY 2720 ATGGAAGATTCACATTCGACATTCGACATTCATCAGCGCATCA---GCTGCATGATCTGC 2776
DB 104 ATGGAAGCCACCCCTGACACTGGAAGACTTGTCTCGCGCTCATTTCCAGCAGACTTGC 163
QY 2777 ACAGACCTTTACGCCACATCTCTAGAGTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2836
DB 164 CGAGGCCCTTTACGCCCTCACCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 223
QY 2837 TTGCARACTGTGTGAGAGACTGTGTGCACACCAATCTGTTTATAAGTTAGGAGTGTCAACAG 2896
DB 224 TCGCAAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 283
QY 2897 GAATTAACATCTCATCAGATGAACAGAAAGGAACATTTTGAAGCCCTGAATTTCTGGAATGT 2956
DB 284 GGACAGGCCGCTGTATGCTCCCAACAGAGCAGGAAACTACGAGACCTTGA----- 331
QY 2957 TTTGAGCAAAATGGAATGCAATTCAGTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3010
DB 332 GCTCGGCAAGGATGGAACACCTCTGAGTTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 391
QY 3011 TTTTATCAGCAGGAAACAGCATAGCTCAATCTTGGACCCAGAGGCAAGGTTAAATTTGTTTC 3070
DB 392 ACGGTTTCCATGTGTATGGTACGGGTGAGCTTCCAGCCCTTCGACCAGTGGCAATGGGAGGCCATTC 451
QY 3071 ATCCCTTGATCGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3130
DB 452 ATCCGTTGGTAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 511
QY 3131 TTACTACAATCAGTAACGATTAACGATATATG 3162
DB 512 TCACCTGTAATGTCACAGAACCGATACATG 543
RESULT 13
BZ411504/c 849 bp DNA linear GSS 04-DEC-2002
LOCUS OGACC43TC ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0027112,
DEFINITION genomic survey sequence.
ACCESSION BZ411504.1 GI:26044995
VERSION BZ411504.1
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 849)
White, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGACC43TM
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..849
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0027112"
FEATURES
source
Location/Qualifiers
1..849
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0027112"
/note="Vector: pBCSK-; Site 1: HindII; 0.7-1.5 kb
methylation filtered genomic DNA library"
```



KEYWORDS  
SOURCE Zea mays  
ORGANISM Zea mays

REFERENCE  
AUTHORS Rabinowicz, P.D., O'Shaughnessy, A.L., Bailly, V., Dedhia, N.,  
Katzenvig, F., King, L., Miller, B., Muller, S., Nascimben, L.,  
Zutavern, T., McCombie, W.R. and Martienssen, R.A.  
Genomic shotgun sequences from Zea mays (methyl-filtered)  
Unpublished (2002)

TITLE  
JOURNAL  
COMMENT Contact: W. Richard McCombie  
Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884  
Fax: 516 367 8874  
Email: mcombie@cshl.org  
Plate: ht52 row: d column: 11  
Seq primer: -21M13UnivFwd  
Class: shotgun  
High quality sequence stop: 635.

FEATURES  
source  
1..635  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="ht52d11"  
/lab\_host="JM107 or DH5a"  
/clone\_lib="WGS-ZmaysF (JM107 adapted methyl filtered)"  
/notes="Organ: immature ears; Site\_1: Xba I; Site\_2: Xba I;  
The vector was digested with XbaI and one nucleotide was  
added by fill in in the recessive 3' end. The genomic DNA  
was nebulized, end repaired, adaptor ligated and size  
fractionated using sephadex. The resulting fragments were  
between 0.8 and 3 kb and were cloned into the vector  
(x/y reads in M13mp19, b/g reads in pUC19). The same  
ligation was transformed in either JM107 or DH5a."

ORIGIN  
Query Match 4.4%; Score 139.4; DB 28; Length 635;  
Best Local Similarity 65.6%; Pred. No. 5.6e-23;  
Matches 254; Conservative 0; Mismatches 121; Indels 12; Gaps 3;

QY 1684 CAGATGGAACCGAAACTCTGTCTGAGTCACCTCGCAAGGTTTCTCCAGCTGAGCAT 1743  
Db 383 CTGCTTGTATACAAATATTGTACCCCTCATCTCGTTACCAAGGTTTCTCTTGTAAACAA 324  
QY 1744 GATATCCAAATTATGTCTGACCTTCATGACGAGTCTACCCCAAGAGAAAGAGCAA 1803  
Db 323 AAGTCCAGATGTGTCTGCTCCACAGAGAAATATACCAAGAAATAGA-----AA 270  
QY 1804 AAACCTGAAGTACTCTGTAAACAGACCATGATGATGACATCCCATGGATATGTT 1863  
Db 269 GGAAGCAGAAGGTGCAACGAGAACAGATGTATAGCACTTCCCATGGACATGTT 210  
QY 1864 GAATCTGATCTAAACACAGCATGAGCGAGCTTATGATGAGACTGATGTTCTGAC 1923  
Db 209 GAATCTTCCAGAAATCAGCAGAGACAGTGTATGATGATGATGATGATGATGATG 150  
QY 1924 ATCAACCGTATTCATCCAGACAACTGTGTATGATGATGATGATGATGATGATG 1983  
Db 149 AACTATCATACTCAACCCAGGTA---GCTCAAGTTGATGTCTGCTGATTTGCAGCCAAG 93  
QY 1984 GATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2043  
Db 92 GGTTGTTCCCAACCAATGATCAATGATGATGATGATGATGATGATGATGATGATG 9  
QY 2044 TCCCAAGATACACAGAGGAGTTACAG 2070  
Db 35 TCGGAAGTAAAGCAGAGTCCCTTACAG 9

Search completed: May 17, 2004, 01:43:09  
Job time : 7754 secs

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: May 17, 2004, 01:47:00 ; Search time 10018 Seconds  
(without alignments)  
4573.127 Million cell updates/sec

Title: US-09-721-114-2  
Perfect score: 5526  
Sequence: 1 MEIVAVDQSGARVVGTCML.....NKVPADFTTISNDNEYMDYR 1057

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgr2 1/USPTO\_spool/US09721114/runat\_14052004\_092251\_4115/app\_query\_faasta\_1.1223  
-DB=GenEmbl -FAST=fastap -SUFFIX=5ge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09721114 @CGN 1 1 5626 @runat\_14052004\_092251\_4115 -NCPU=6 -ICPU=3  
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_on.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rtd.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	5526	100.0	4310	6	BD083699	Novel gen
2	5516	99.8	3896	8	AF326768	Oryza sat
3	5191	93.9	9455	6	BD083700	Novel gen
4	5191	93.9	150594	8	AP001859	Oryza sat
5	5179	93.7	3626	8	AK103364	Oryza sat
6	393	7.1	3291	6	AX590041	Sequence
7	298	5.4	8648	6	AR368885	Sequence
8	298	5.4	8648	8	AF319968	Arabidops
9	298	5.4	17341	6	AR368884	Arabidops
10	298	5.4	101665	8	ATP15N18	Arabidops
11	241.5	4.4	7036	10	MUSCYCLOA	Mus muscu
12	234.5	4.2	32367	10	AF520421S2	Danio rer
13	234.5	4.2	22782	2	EX842684	Drosophil
14	233	4.2	184271	3	AC009256	Drosophil
15	233	4.2	198961	3	AC010919	Drosophil
16	233	4.2	323844	3	AE003501	Drosophil
17	227.5	4.1	212955	2	AC136676	Rattus no
18	225	4.1	160827	3	AC099025	Drosophil
19	223.5	4.0	193362	3	AC007579	Drosophil
20	223.5	4.0	260367	3	AE003808	Drosophil
21	223	4.0	3114	9	HSATRX10	Homo sapien
22	223	4.0	4808	3	AY075323	Drosophil
23	223	4.0	7479	9	AB102641	Homo sapi
24	223	4.0	10216	9	HSU72938	Homo sapien
25	223	4.0	10330	6	AX828379	Sequence
26	223	4.0	10330	9	HSU72937	Homo sapien
27	223	4.0	10452	6	AX336451	Sequence
28	223	4.0	10452	6	AX828404	Sequence
29	223	4.0	10452	9	HSU72936	Homo sapien
30	223	4.0	47323	9	AC005937	Homo sapi
31	223	4.0	84474	9	AL662854	Human DNA
32	223	4.0	103780	9	AL773541	Human DNA
33	223	4.0	110000	2	BX276116_03	Continuation (4 of
34	223	4.0	135982	9	HS34606	Human DNA s
35	223	4.0	153440	3	AC105263	Drosophil
36	223	4.0	156272	9	AL669830	Human DNA
37	223	4.0	178688	2	AL713893	Homo sapi
38	223	4.0	276131	3	AE003543	Drosophil
39	222	4.0	10652	10	AF441786	Mus muscu
40	221	4.0	3829	9	HSXNP08	Homo sapi
41	221	4.0	6542	9	AB014524	Homo sapi
42	221	4.0	6655	9	AY099469	Homo sapi
43	221	4.0	10358	9	HSU75653	Human zinc
44	221	4.0	151865	2	AC025197	Homo sapi
45	221	4.0	347050	3	PFA929351	Plasmodiu

ALIGNMENTS

RESULT 1

BD083699  
LOCUS BD083699 4310 bp DNA linear PAT 27-AUG-2002  
DEFINITION Novel gene participating in response to brassinosteroid.  
ACCESSION BD083699  
VERSION BD083699.1 GI:22629109  
KEYWORDS JP 2001327287-A/1.  
SOURCE Oryza sativa  
ORGANISM Oryza sativa

REFERENCE  
AUTHORS Hirochika, H., Yamazaki, T. and Miyao, A.  
TITLE Novel gene participating in response to brassinosteroid  
JOURNAL Patent: JP 2001327287-A 1 27-NOV-2001;  
NATL INST OF AGRICULTURAL RESOURCES, BIO ORIENTED TECHNOLOGY  
RESEARCH ADVANCE JAPAN FORAGE SEED ASSOCIATION  
COMMENT OS Oryza sativa (rice)  
PN JP 2001327287-A/1  
PD 27-NOV-2001  
PF 19-MAY-2000 JP 2000149106  
PI HIROCHIKA HIROCHIKA, TOKIO YAMAZAKI, AKIO MIYAO  
PC C12N15/09, C12N15/00  
CC Novel gene participating in response to brassinosteroid FH  
Key Location/Qualifiers  
FT CDS Location/Qualifiers  
1..4310  
/organism="Oryza sativa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4530"

ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 4310  
Score: 5526.00 Matches: 1057  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-09-721-114-2 (1-1057) x BD083699 (1-4310)

QY 1 MetGluIleValAlaValAspGlnGluGlyAlaArgValGlyThrAsnCysMetLeu 20  
DB 655 ATGGAGATTGTTGCAGTAGATCAGGAGGAGCTCGTGTGTTGGGACCACTGTATGCTT 714  
QY 21 AlaArgGlyGlyThrGlyAlaValAlaProValLeuGluLeuThrAlaThrProArgGln 40  
DB 715 GCTCGTGGTGGAACTGCTGTAGCCGCGAGTGTGGAGCTGACGAGCGCTCGTCAG 774  
QY 41 AspAlaAlaAlaGluAlaGlyValAspGluProAlaGlnHisGlnCysGluHisPheSer 60  
DB 775 GATGCGAGCGCTGAGCTGTGTAGCGAACCGGCACACACCAATGCGAGCATTTCTCC 834  
QY 61 IleArgGlyTyxValAlaLeuLeuGlnLysAspProLysPheCysSerLeuSerArg 80  
DB 835 ATAAGAGGGTATGTGCTCTTTTCAGAGAGAGGATCCAAAATTCGTCTCTATCTCGG 894  
QY 81 IlePheHisAspGlnLysCysAspGluHisLysAlaSerSerProPheSerVal 100  
DB 895 ATTTTCCATGACCAAGAAAATGTGATGAACACAAAGCTAGTTCAAGCCATTTCTGTA 954  
QY 101 AlaLysPheArgArgTrpAspCysSerLysCysLeuAspLysLeuLysThrSerAspAsn 120  
DB 955 GCAAAAGTTTCGACGATGGATGCTCGAAGTCTTGGATTAAGTTGAAAACCTTCAGATAAT 1014  
QY 121 GlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIle 140  
DB 1015 GGAAACGACCCAGAACTCTTCCGCAACGAGATGGCAGAGTGATGGTTGCTCCATC 1074  
QY 141 ThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysValSerProSer 160  
DB 1075 ACATTTGTTGCGAGCACCTTTTGTGCTGCTAGTGTGGTTCCTCCAAAAGTGTCTCTAGC 1134  
QY 161 ThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGlnGlu 180  
DB 1135 ACACATCATCTCAAGGGAAGATGCTGATAGATCAACTCTTCCAAAAGAGTGTCAAGAA 1194  
QY 181 GlyAsnAspSerLysCysAsnAlaProSerClyLysAsnGlyAlaAlaGluAlaAsnThr 200  
DB 1195 GGCATGACTCCAAATGCAATGCCCTTCTGGCAAGATGAGCTGTGAGCCCAATACT 1254  
QY 201 AspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyraAspValAlaAlaAsnVal 220  
DB 1255 GATTCAACCAATGAAGATTTGCAAGGCCAGCCCAAAATATATGATGTGGCAGCAATGTC 1314  
QY 221 SerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIleThrTrp 240  
DB 1315 TCTGAGGACACACTTCTGTTGATGTTGGGGCTTTTACCTGAAGTTCCTCCAGATTACATGG 1374  
QY 241 HisIleGluValAsnGlnValAspGlnProProSerThrProLysLeuSerGluValVal 260  
DB 1375 CACATGAGATTAATGCTGCAGATCAACTCCATCCACTCCAAACCTTCTGAAGTGGTC 1434  
QY 261 LeuLysArgAsnGluAspGluAsnGlyLysThrGluGluThrLeuValAlaGluGlnCys 280  
DB 1435 CTCAAAAGAAATGAAGATGAAAATGGAAAACCTGAAGAGACTCTTGTGTGTCGACGATGC 1494  
QY 281 AsnLeuThrLysAspProAsnProMetSerGlyLysGluArgAspGlnValAlaGluGln 300  
DB 1495 AATTGACCAAGATCTTAACCCCAATGCTCGAAGGAACCTGATCAGTTCTGAGCAG 1554  
QY 301 CysAsnLeuThrLysAspProLysProValSerGlyGlnLysCysGluGlnIleCysAsn 320  
DB 1555 TGCATTTGACCAAGATCGAAACCACTGCTGGCAGAAATGTGACGACAGATCTGCAAT 1614  
QY 321 GluProCysGluGluValValLeuLysArgSerSerLysSerLysArgLysThrAspLys 340  
DB 1615 GAGCATGTGAGAAAGTGTCTCAAAAGAAAGCTCCAAATCTAAGAGGAAGACCGATTAAG 1674  
QY 341 LysLeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAsp 360  
DB 1675 AAGTTGATGAAGAGCAGCAGCAGCAAGAAACGCACTGCCAGGCTGATGTTTCAGAT 1734  
QY 361 AlaLysLeuCysArgArgLysProLysValArgLeuLeuSerGluIleIleAsnAla 380  
DB 1735 GCAAAAGCTTTTTCGAGAAAGCCCAAAAAGGTGCGGCTTCTATCAGAAATATATAATGCT 1794  
QY 381 AsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspProCys 400  
DB 1795 AACCAAGTTGAGGATTTAGAGTGAAGAGTTCATGCTGAAAATGCGCTGATCCCTGT 1854  
QY 401 GluAspAspArgSerThrIleProValProMetGluValSerMetAspIleProValSer 420  
DB 1855 GAGGATGATAGAGTACCATCCCGTCCGATGGAAGTAAAGTATATTCCTGTTAGC 1914  
QY 421 AsnHisThrValGlyLysGluAsnGlyLeuLysSerSerLysAsnLysThrLysArgLysTrp 440  
DB 1915 AACCTACAGTGGGAGAGATGGGTTAAATCAAGTAAGAACACACAAACGCAATATC 1974  
QY 441 SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysLysArg 460  
DB 1975 TCTGATGTTGTAGATGATGATGATCATCTTATGAATGCTGATGGAATGAAAAGAAAGA 2034  
QY 461 ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
DB 2035 ACTGGAAGTGGCATCACACAGTTGCTCATCCAGCTGGAAATTTGACAAACAAAAGTGT 2094  
QY 481 ThrProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAspThr 500  
DB 2095 ACACCACTGCGAGTACTCAGCATGATGATGAGATGATGATGATGATGATGATGATGAT 2154  
QY 501 AsnMetHisLysThrAspValCysGlnHisValSerGluIleSerThrGlnArgCysSer 520  
DB 2155 AATATGATGAAGACAGATGCTGTCAGCATGATGATGATGATGATGATGATGATGATGAT 2214

QY 521 SerIysGlyLysThrAlaGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLys 540  
DB 2215 TCAAGGGGAACACAGGGGTTTGTAGTAAGGGGAACACATTCAGCTGTAGTACCAAA 2274  
QY 541 TyrGlyGlyLysThrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGln 560  
DB 2275 TATGGTGGTGAAGACACAGAAATGGTTCAGAAATATATCTACTCAGCGCAGAGATCAA 2334  
QY 561 CysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLysValSerProAlaGlu 580  
DB 2335 TGCAGATGGAAACCGAANAATCTGTCTGAGTCACTCGGCAAGGTTTCTCAGCTGAG 2394  
QY 581 HisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLys 600  
DB 2395 CATGATATCCAAATATGTCTGACCTTCATGACGAGAGTCTACCCAGGAAGAAAAGAAG 2454  
QY 601 GlnLysLeuGluValThrArgGlyLysGlnThrMetIleAspAspIleProMetAspIle 620  
DB 2455 CAATAACTTGAAGTACTCTGTGAATAACACCATGATGATGATGATGATGATGATGATG 2514  
QY 621 ValGluLeuLeuAlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSer 640  
DB 2515 GTTGAATCTGTAGTAAACACGATGAGAGGAGCTATGATGATGATGATGATGATGAT 2574  
QY 641 AspIleAsnArgIleGlnSerLysThrAlaAspAspAspCysValIleValAlaAla 660  
DB 2575 GACATCAACCGTATTCATCCAAAGACACTGCTGATGATGATGATGATGATGATGATG 2634  
QY 661 LysAspGlySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGlnLysSerLeu 680  
DB 2635 AAGGATGTTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2694  
QY 681 AlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGluSer 700  
DB 2695 GCATCCCAAGTACACAGAGGAGTACAGGGTCTATTGGCATTGACCAACAGAGTCT 2754  
QY 701 ProHisProGlnAsnPheGlnSerThrGlnGlnGlnGlnThrHisLeuArgMetGluGlu 720  
DB 2755 CCACATCTCAGAACTTTTCTGATCTACTCAGAAACAGCAGACATTTGCGATGGAGAA 2814  
QY 721 MetValThrIleAlaAlaSerSerProLeuPheSerHisHisAspAspGlnTyrIleAla 740  
DB 2815 ATGGTCACTATTGCTGCAAGCTCACCACCTATTTCATCATCATCATCATCATCATCAT 2874  
QY 741 GluAlaProThrGluHisTyrGlyArgLysAspAlaLysIleLysLeuThrTyrGluGlnPhe 760  
DB 2875 GAAGCAACCACTGAACATTTGGGCGCGTAAGGACGCAAGAGCTAACGTTGGGAGCAATT 2934  
QY 761 LysAlaThrThrArgAsnSerProAlaAlaThrCysGlyValAlaGlnPheArgProGlyIle 780  
DB 2935 AAGGCCACTACAGAAATTTCTCAGCAGCAACATGTTGTTCTCAATTTAGACCTGGTATC 2994  
QY 781 GluAlaValAspLeuThrSerThrHisValMetGlySerSerSerSerSerSerSerSer 800  
DB 2995 CAAGCAGTTGACTTGACTTCTTCTACTCATGTCTATGGGATCTTCCAGCAATTTATGCACT 3054  
QY 801 GlnProValIleAlaProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAla 820  
DB 3055 CAACCAAGTAAATGGGCACTGGACCGCTATGCTGAAAGAGCGGTTAACCAAGTTCATGCA 3114  
QY 821 ArgAsnPheProSerThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsn 840  
DB 3115 AGAAATTTTCCAGCAACATAGCAACCATCGAAGCGAGTAAGTTATGTATGTCATGCGAGAA 3174  
QY 841 AlaGlyGlnValValLeuTyrProLysGluSerMetProAlaThrHisLeuLeuArgMet 860  
DB 3175 GCTGGACACAGTAGTCTTGTATCTCTAAGAAATCATGCTTCGACGAGCATCTTCTGAAAG 3234  
QY 861 MetAspProSerThrLeuAlaSerPheProAsnTyrGlyThrSerSerArgAsnGlnMet 880  
DB 3235 ATGGATCCATCAACATTAGCAAGCTTCCCACTATGAACTTCTAGCAGAACCAACAGATG 3294

QY 881 GluSerGlnLeuHisAsnSerGlnTyrAlaHisAsnGlnTyrLysGlySerThrSerThr 900  
DB 3295 GAGTCTCAACTTTCATAATTTCTCAGTATGCACATAATCAGTACAAAGGATCAACCCAGCA 3354  
QY 901 SerTyrGlySerAsnLeuAsnGlyLysIleProLeuThrPheGluAspLeuSerArgHis 920  
DB 3355 TCATATGGCAGTAACCTGAATGGAAGATTCATTCAGCATTCGAAGACTTATCAGCGCAT 3414  
QY 921 GlnLeuHisAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlySer 940  
DB 3415 CAGCTGCATGATCTGCACAGACCTTTACGCCACATCTTAGAGTTGGTGTGCTGCTCC 3474  
QY 941 LeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLys 960  
DB 3475 TTGCTGCAGAGGAATTCGAACTGGTGGAGAACTGGGCACACAACTCTGGTTATAAG 3534  
QY 961 LeuGlyValSerThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAla 980  
DB 3535 TTAGAGTGTCAACAGGAATACATCGCATCAGATGAACAGAAAGAAACATTTTGAAGCC 3594  
QY 981 LeuAsnSerGlyMetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSer 1000  
DB 3595 CTGAATTTCTGGAATTTTTCAGCAAAATGGATGCAATTCGATTCGCTGTCTGTAGCTCC 3654  
QY 1001 SerAlaAspPheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGly 1020  
DB 3655 AGTGCAGATTTTTCAGCGAGGAAACAGCATGCTCAATCTTGACAGAGCAAGGT 3714  
QY 1021 LysMetValHisProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAsn 1040  
DB 3715 AAAATGGTTCATCCCTTGGATCGTTCGTCAGACAGGATATCTGTATTAACCAAGAAC 3774  
QY 1041 ProAlaAspPheThrThrIleSerAsnAspAsnGluTyrMetAspTyrArg 1057  
DB 3775 CCAGCTGATTTTACTACATCAGTAACAGATACAGATATATGATGATACCGC 3825

RESULT 2  
AF326768 3896 bp mRNA linear PLN 05-SEP-2001  
LOCUS Oryza sativa embryonic flower 1-like protein mRNA, complete cds.  
DEFINITION AF326768  
ACCESSION AF326768  
VERSION AF326768.1 GI:15430698  
SOURCE Oryza sativa  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
REFERENCE 1 (bases 1 to 3896)  
AUTHORS Aubert,D., Chen,L., Moon,Y.H., Martin,D., Castle,L.A., Yang,C.H.  
and Sung,Z.R.  
TITLE EMP1, a novel protein involved in the control of shoot architecture  
and flowering in Arabidopsis  
JOURNAL Plant Cell 13 (8), 1865-1875 (2001)  
MEDLINE 21380417  
PUBMED 11487698  
REFERENCE 2 (bases 1 to 3896)  
AUTHORS Moon,Y.-H., Chen,L. and Sung,Z.R.  
TITLE Direct Submission  
JOURNAL Submitted (06-DEC-2000) Plant and Microbial Biol. Dept., University  
of California, Berkeley 361 Koshland Hall, Berkeley, CA 94720, USA

FEATURES  
source  
1..3896  
/organism="Oryza sativa"  
/mol\_type="mRNA"  
/db\_xref="taxon:4530"  
241..3414  
/note="EMF1"  
/codon\_start=1  
/product="embryonic flower 1-like protein"  
/protein\_id="AAK98529.1"  
/db\_xref="GI:15430698"  
/translation="MEIVAVDQEGARVVGTNCMLARGGTGAVPVLTLTATPRQDAAA"

CDS

EAGVDEPAHQCEHSIRGYVALLQKDPKFCISLSRIIFHDQKKDBHKASSPFSVAK  
FRWDCSLDKLTKTSDNGTAPRTLPKQNGTSDGCSITFVRSTFPVPSVGSQKVS  
TOSQGNADSLTKSVOEGNDKCNAPSGKNGAEANTDTPKQKLOGPAQNDVAA  
NVSSENTSVUGALREVPQITWHIEVNGADQPPSTPKLSEVVLKRENDENKRETLV  
AECNLTKDPNPKSGBRQDVAQCNLTDPKPSGQCEQICNEPCEVSEVWIKSSKS  
REKTDKLAKKQKSKRTAQADVSDAKLCRRKPKKVELLSIIVNANOVSRSDEHV  
RENAADPCEDDRSTTIPVPEVSMOIPVSNHTVGEDGLKSXKNTKRYSDVWHDGSSL  
NMLNGKKRTGSHVHTVAHPAGNLSNKKVTPASTQHDNDENDTENGJLDVNNKTDVC  
QHSVISTORCSSKKTAGLSKGTSHSAASTKYGCESTRNGONTHVLSAEDOCOWETE  
NSVLSHSAKVSABEDIOIMSDLHESQSLPKKKXKOKLEVREKOTMIDDIIPMDIVELL  
AKNQHOLATWETDCSDINRQSKTADDCVIVAAKXGSIYASVFTNVSQKSLAS  
QSKELQGLHALITQESHPQNFQSTQEQTHLREEMVTIAAASPUSFHHQOYIA  
EAPTHWRKQKLUKTQFQKATNSPAATCGAFREGIOAVDLTSTPHVMSGSSNYA  
SRQVIALPDLYBAEAVNOVHARNFPTIATMEASKLCDRENAGOVVLYPKESPATH  
LILRMDPSTLASFPNYGTSRRNOMESQHLNSQYAHNQYKGSTSYSGSNLNGKIPLTF  
EDLSRQLHLHLRPLRPRPGVGLSLLOKEIANKSENGCTQSGYKLCVSTGITSOM  
NEKHEFEALNSWPSAKNALQLGVSSSAYFLSARNSIAQSWTEGKSKVHPLDRFV  
RQDICITNKNPADFTTISNDNEYMYR"

ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 3896  
Score: 5516.00 Matches: 1055  
Percent Similarity: 99.91% Conservative: 1  
Best Local Similarity: 99.81% Mismatches: 1  
Query Match: 99.82% Indels: 0  
DB: 8 Gaps: 0

US-09-721-114-2 (1-1057) x AF326768 (1-3896)

Qy	1	MetGluIleValAlaValAspGlnGluGlyAlaArgValValGlyThrAsnCysMetLeu	20
Db	241	ATGGAGATTGTTGCGATGATGTCGAGAGGAGCTGCTGTTGTTGGACGACACTGATGTT	300
Qy	21	AlaArgGlyGlyThrGlyAlaValAlaProValLeuGluLeuThrAlaThrProArgGln	40
Db	301	GCTCGTGTGGTGAACCTGCTCTAGCGCCAGTGTGGAGCTGACACGACGCGCTCGTCAG	360
Qy	41	AspAlaAlaAlaGluAlaGlyValAspGluProAlaGlnHisGlnCysGluHisPheSer	60
Db	361	GATGACGCGCTGAAGCTGGTGTAGACGAAACCGGACAAACCAATGCGAGATTCTCC	420
Qy	61	IleArgGlyTyrValAlaLeuLeuGlnIlyValysAspProLyPheCysSerLeuSerArg	80
Db	421	ATAAGAGGATGTTGCTCTCTTTCAGAGAAGAGGATCCAAATTTCTGCTCTCTATCTCG	480
Qy	81	IlePheHisAspGlnIlyLysCysAspGluHisLysAlaSerSerProPheSerVal	100
Db	481	ATTTTCCATGACCCAGAAAATGTGATGAACACAAAGCTAGTTCAGGCCCATTTTCTGTA	540
Qy	101	AlaLysPheArgArgTyrAspCysSerLysCysLeuAspLysLeuLysThrSerAspAsn	120
Db	541	GCAAGTTTCAGCATGGATGCTCGAAGTCTCTGGATAAGTTCAAAACITTCAGATAAT	600
Qy	121	GlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIle	140
Db	601	GGAAACAGCACCAAGAACTTCCCGCAAGACAGAAATGGCACAAAGTATGTTGTCTCCATC	660
Qy	141	ThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysValSerProSer	160
Db	661	ACATTTGTTCCGAGCACTTTTGCCCTGCTAGTGTGTTTCCCAAAAAGTGTCTCTAGC	720
Qy	161	ThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGlnGlu	180
Db	721	ACACAATCATCTCAAGGGAAGAATGCTGATAGATCAACTCTTCCAAAGAGTGTGCAAGAA	780
Qy	181	GlyAsnAspSerLysCysAsnAlaProSerGlyLysAsnGlyValaAlaGluAlaAsnThr	200
Db	781	GGCAATGACTCCAAATGCAATGCGCTTCTGCGAAGATGGAGCTGCTGAGGCCCAATACT	840
Qy	201	AspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAlaAlaAsnVal	220
Db	841	GATTCCACCAATGAAGATTTGCAAGGGCCAGCCCAAAATTTATGATGTGGCAGCAATGTC	900
Qy	221	SerGluAspAsnThrSerValAspValGlyValAlaLeuProGluValProGlnIleThrTyr	240
Db	901	TCTGAGGACACACTTCTGTTGATTTGGGCTTTACTGAAGTTCCTCCAGATTACATGG	960
Qy	241	HisIleGluValAsnGlyAlaAspGlnProProSerThrProLysLeuSerGluValVal	260
Db	961	CACATAGAAGTAATAGTGTGAGATCAACCTCCATCCATCCATCCAAAACTTTCTGAAGTGTG	1020
Qy	261	LeuLysArgAsnGluAspGluAsnGlyLysThrGluGluThrLeuValAlaGluGlnCys	280
Db	1021	CTCAAAAGATGAAGATGAATGAATGAAGAAACTGAGAGACTCTTGTGTGTCGACAGTGC	1080
Qy	281	AsnLeuThrLysAspProAsnProMetSerGlyLysGluArgAspGlnValAlaGluGln	300
Db	1081	AATTGACCAAGATCTTAACCCCAATGCTCGAAAAGGAAGTGCATCAGGTGCTGAGCAG	1140
Qy	301	CysAsnLeuThrLysAspProLysProValSerGlyClnLysCysGluGlnIleCysAsn	320
Db	1141	TGCATTTTGACCAAGATCCGAAACCCAGTGTCTGGCGAAGATGTGAGCAGATCTGCAT	1200
Qy	321	GluProCysGluGluValValLeuLysArgSerSerLysSerLysArgLysThrAspLys	340
Db	1201	GAGCCATGTGAAGAGATTGTTCTCAAAAAGAAAGCTCCAAATCTAAGAGGAAGACGATAAG	1260
Qy	341	LysLeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAsp	360
Db	1261	AAGTTGATGAAGAGACGACGACAGCAAGAAACGCACTGCCAGCTGATGTTTCAGAT	1320
Qy	361	AlaLysLeuCysArgArgLysProLyLysValArgLeuLeuSerGluIleIleAsnAla	380
Db	1321	GCAAGCTTTGCGAGAGAAACCCAAAGAGTGGCGCTTCTATCAGAAATTTGTAATGCT	1380
Qy	381	AsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspProCys	400
Db	1381	AACAGATTGAGGATCTAGAAGTGAAGAGTTCATCGTGAATGCGCTGATCCCTGT	1440
Qy	401	GluAspAspArgSerThrIleProValProMetGluValSerMetAspIleProValSer	420
Db	1441	GAGGATGATAGAGTACCATCCCGTCCCGATGGAAGTGAAGTGAATTCCTGTGAGC	1500
Qy	421	AsnHisThrValGlyGluAspGlyLeuLysSerSerLysAsnLysThrLysArgLysTyr	440
Db	1501	AACCATACAGTGGGAGAGATGGGTTTAAATCAAGTGAAGAACCAAGCAAAACCGCAATAC	1560
Qy	441	SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysLysArg	460
Db	1561	TCTGATGTTGTAGATGATGGATCATCACTTATGACTGGCTGAATGCAAAAAGAAAGA	1620
Qy	461	ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal	480
Db	1621	ACTGGAAGTGTGCATCACACAGTTGCTCATCCAGCTGGGAAATTTGAGCAACAAAAAATG	1680
Qy	481	ThrProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAspThr	500
Db	1681	ACCCCACTGGAGTACTCAGCATGATGATGATGATGATGATGATGATGATGATGATGATG	1740
Qy	501	AsnMetHisLysThrAspValCysGlnHisValSerGluLysSerThrGlnArgCysSer	520
Db	1741	AATATGATGAAGACAGATGCTGTGACAGCTGATCAGAAATCTCCACACAGAGGTGCTCA	1800
Qy	521	SerLysGlyLysThrAlaGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLys	540
Db	1801	TCAAGGGGAAAACAGCGGGTTTGAGTAAGGGGAAAACACATTCAGCTGCTAGTACCAAA	1860
Qy	541	TyrGlyGlyGluSerThrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGln	560
Db	1861	TATGTTGTTGAAGACCCAGAAATGGTGCAGAACATACATGTTACTCAGCGCAGAGATCAA	1920
Qy	561	CysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLysValSerProAlaGlu	580
Db	1921	TGCCAGATGGAACCGGAAACTCTGTTCTGAGTCACTCGGCAAGGTTTCTCCAGCTGAG	1980

QY 581 HisAspIleGlnIleMetSerAspLeuHisGlnSerLeuProIlySerLysLys 600  
DB 1981 CATGATATCCAAATATGTCGACCTTCATGAGCAGAGTCTACCCAGAGAAAGAG 2040  
QY 601 GlnIlySerGlnValThrArgGluLysGlnThrMetIleAspIleProMetAspIle 620  
DB 2041 CAAAACTTCAGAGTCTCGTGAAGAAACAGACCATGATAGATGACATCCCATGGATAT 2100  
QY 621 ValGlnLeuLeuAlaLysAsnGlnHisGlnArgGlnLeuMetThrGlnThrAspCysSer 640  
DB 2101 GTTGAACTGTAGCTAAACACAGCATGAGAGCAGCTTATGACTGAGAGTATTGTTCT 2160  
QY 641 AspIleAsnArgIleGlnSerLysThrAlaAspAspCysValIleValAlaAla 660  
DB 2161 GACATCAACGTTATTCATCAAGACAACTGCTGATGATGATGTTGTAATAGTCTGCC 2220  
QY 661 LysAspGlySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGlnLysSerLeu 680  
DB 2221 AAGGATGGTTCAGATTATGATCATCAAGTGTGTTGACACTAATTCACACAGAGTCTCTG 2280  
QY 681 AlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGluSer 700  
DB 2281 GCATCCCAAGTACACAGAGAGGATTACAGGGTCAATTTGGCATTGACACACAGAGTCT 2340  
QY 701 ProHisProGlnAsnPheGlnSerThrGlnGlnGlnGlnThrHisLeuArgMetGluGlu 720  
DB 2341 CCATCCTCAGAACTTTCAGTCTACTCAGACAGACAGACATTTGCCGATGGAGAA 2400  
QY 721 MetValThrIleAlaAlaSerSerProLeuPheSerHisHisAspAspGlnTyrIleAla 740  
DB 2401 ATGGTCACTATTGCTGACGCTCACCACCTATTTCATCATCATGATCATCATGATGCT 2460  
QY 741 GluAlaProThrGlnHisIleThrGlyArgLysAspAlaLysLysLeuThrTyrGlnPhe 760  
DB 2461 GAAGACCAACTGAACATTGGGCGGTPAAGAGCGCAAGAGCTAACGTGGGAGCAATTT 2520  
QY 761 LysAlaThrThrArgAsnSerProAlaAlaThrCysGlyAlaGlnPheArgProGlyIle 780  
DB 2521 AAGGCCACTACAGAAATTCCTCAGCAGCAACATGTTGCTGCTCAATTAGACCTGGTATC 2580  
QY 781 GlnAlaValAspLeuThrSerThrHisValMetGlySerSerSerAsnTyrAlaSerArg 800  
DB 2581 CAAGCAGTTGACTTGACTTCTACTCATGTCATGGGATCTTCAGCAATATGATCATCTGC 2640  
QY 801 GlnProValIleAlaProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAla 820  
DB 2641 CAACCAAGTAAATGGCCACTGGACCGTATGCTGMAAGACGGTTAACCAAGTCCATGCA 2700  
QY 821 ArgAsnPheProSerThrIleAlaThrMetGluAlaSerLysLeuCysAspArgAsn 840  
DB 2701 AGAAATTTCCAGCACATAGCAACCATGGAAGCGAGTAAGTTATGTGATCGAGAAAT 2760  
QY 841 AlaGlyGlnValValLeuTyrProLysGluSerMetProAlaThrHisLeuLeuArgMet 860  
DB 2761 CTGGGACAGTAGTCTTTGTATCTCTAAAGAAATCCATGCTCGACGCGATCTCTGAGATG 2820  
QY 861 MetAspProSerThrLeuAlaSerPheProAsnTyrGlyThrSerSerArgAsnGlnMet 880  
DB 2821 ATGGATCCATCAACATTAGCAAGCTTCCCACTATGGAACCTCTAGCAGAACCATG 2880  
QY 881 GluSerGlnLeuHisAsnSerGlnTyrAlaHisAsnGlnTyrLysGlySerThrSerThr 900  
DB 2881 GAGTCTCAACTTCAATAATCTCAGTAGTCACATAATCAGTACAAAGCATCAACACGACCA 2940  
QY 901 SerTyrGlySerAsnLeuAsnGlyLysIleProLeuThrPheGluAspLeuSerArgHis 920  
DB 2941 TCATATGGCAGTAACCTGAATGGAAAGATTCCATTGACATTCGAGACTTATCAGCGCAT 3000  
QY 921 GlnLeuHisAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlySer 940  
DB 3001 CAGCTGCATGATCTGCACAGACCTTTACGCCACATCTCAGAGTTGGTGTGCTGCTGCC 3060  
QY 941 LeuLeuGlnLysGluIleAlaAsnTyrSerGluAsnCysGlyThrGlnSerGlyTyrLys 960

DB 3061 TTGTCGACAGAGAAATTCAAACTGGTCGAGAACTGGGCACACAAATCTGTTTATAAG 3120  
QY 961 LeuGlyValSerThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAla 980  
DB 3121 TTAGAGTGTTCACAGGATAACATCCGATCAGATGACAGAAAGCAATTTTGAAGCC 3180  
QY 981 LeuAsnSerGlyMetPheSerAlaLysTyrAsnAlaLeuGlnLeuGlySerValSer 1000  
DB 3181 CTGAATTTCTGGAAATGTTTTTCAGCAAAATGGAATGCAATTCAGTTGGTCTCTGTAGCTCC 3240  
QY 1001 SerAlaAspPheLeuSerAlaArgAsnSerIleAlaGlnSerTyrThrArgGlyLysGly 1020  
DB 3241 AGTGCAATTTTTTATCAGCAGGAAACAGATAGCTCAATCTTGGACCAAGCAAGGGT 3300  
QY 1021 LysMetValHisProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAsn 1040  
DB 3301 AAAATGGTTTCATCCCTTGCATCGTGTGTGAGACAGGATATCTGTATAACTAACAGAAC 3360  
QY 1041 ProAlaAspPheThrThrIleSerAsnAspAsnGluTyrMetAspTyrArg 1057  
DB 3361 CCAGCTGATTTTACTACAAATCAGTAACGATACGAGTATATGATATACCGC 3411  
RESULT 3  
BD083700 9455 bp DNA linear PAT 27-AUG-2002  
LOCUS Novel gene participating in response to brassinosteroid.  
DEFINITION BD083700  
ACCESSION BD083700.1 GI:22629310  
VERSION JP 2001327287-A/2.  
KEYWORDS Oryza sativa  
SOURCE Oryza sativa  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
REFERENCE 1 (bases 1 to 9455)  
AUTHORS Hirochika,H., Yamazaki,T. and Miyao,A.  
TITLE Novel gene participating in response to brassinosteroid  
JOURNAL Patent: JP 2001327287-A 2 27-NOV-2001;  
NATL INST OF AGRICULTURAL RESOURCES, BIO ORIENTED TECHNOLOGY  
RESEARCH ADVANCEMENT INSTITUTION  
COMMENT OS Oryza sativa (rice)  
PN JP 2001327287-A/2  
PD 27-NOV-2001  
PF 19-MAY-2000 JP 2000149106  
PI HIROHIKO HIROCHIKA,TOKIO YAMAZAKI,AKIO MIYAO  
PC C12N15/09,C12N15/00  
CC Novel gene participating in response to brassinosteroid FH  
Key source Location/Qualifiers  
FT source 1..9455  
FT /organism='Oryza sativa (rice)'.  
FEATURES  
Location/Qualifiers  
1..9455  
/organism='Oryza sativa'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:4530'  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,06e-302 Length: 9455  
Score: 5191.00 Matches: 1055  
Percent Similarity: 68.06% Conservative: 0  
Best Local Similarity: 68.06% Mismatches: 2  
Query Match: 93.94% Indels: 495  
DB: 6 Gaps: 3  
US-09-721-114-2 (1-1057) x BD083700 (1-9455)  
QY 1 MetGluIleValAlaValAspGlnGluGlyAlaArgValValGlyThrAsnCysMetLeu 20  
DB 4223 ATGGAGATTGTTGCAGTAGATCAGAGGAGGAGCTGTTGTTGGGACGAACTGATGCTT 4282  
QY 21 AlaArgGlyGlyThrGlyAlaValAlaProValLeuGluLeuThrAlaThrProArgGln 40



Db 4283 |||||GCTCGTGGTGGAACTGGTCTGTAGCGCCAGTGTGGAGCTGACAGCGCTCGTCAG 4342  
QY 41 AspAlaAlaGluAlaGlyValAspGluProAlaGlnHisGlnCysGluHisPheSer 60  
Db 4343 GATGACGCGCTGAGCTGTGTAGACGAACCGSCACAAACCAATSCGAGCATTTCTCC 4402  
QY 61 Ile----- 61  
Db 4403 AT-AAGGTAATCAATTTCTGTATTTCCAAATCCAGTATCGGTTGGGATGAATATGAA 4461  
QY 61 ----- 61  
Db 4462 TCGGCATGTCATGCCATATTGCACTGTGTCATGGAAGAGTATGATGATCGTGGTTTGTG 4521  
QY 61 ----- 61  
Db 4522 CACAGTTTGTGTGGCACTTATATGATGTCATCTGTGTTTGTAGATCGTATACACTGGGTC 4581  
QY 61 ----- 61  
Db 4582 GACATGCTTATGACTTTTGGTTGCAATTTAGGAAGTCAATACATCCACTACTAGTCTCTATAT 4641  
QY 61 ----- 61  
Db 4642 CTAGCCATGTGAACACTCATTTATAGCCATAGCACAGCTAGCAGGCTAGCAGCAAAAAATATA 4701  
QY 61 ----- 61  
Db 4702 TATAATATTGCAATATATGTTGGTTGTCATGTAATCTTTATATCTCTACGTACATCCATTTA 4761  
QY 61 ----- 61  
Db 4762 ATATCTTCAATGATGAATCTGAGCACAGATGTTGTGAGTGTCTACATATGCAATGTCGT 4821  
QY 61 ----- 61  
Db 4822 ATGTGTGTTCAATGAGTGTGATCATATTTGTTGTTGGGTGGCGATGCAATTTAT 4881  
QY 61 ----- 61  
Db 4882 CAGGCCATGCTGTAGGCTGTAGCTAGATATTGTTGTTGTATATTATTTCTGTTGAACAA 4941  
QY 61 ----- 61  
Db 4942 GCTGATTACTAATGAATAATTAACCTTTTGGGTACACTCATATATTGGGCCCTACATTTT 5001  
QY 61 ----- 61  
Db 5002 TGTAATCAATTTTCTCTGTGCTGAGGTTTCAGCATAAACTTTTATCATAGCATGTT 5061  
QY 61 ----- 61  
Db 5062 TACATCTAGAGATCTTAGAACTGATGGTTTCTTCATATTTGCATTTATGTTGATTTG 5121  
QY 61 ----- 61  
Db 5122 ATAGTCCATTATTTTAAAGCCTTTTCAATTTGTTTAGAGATCTTAGAGATGATATATAT 5181  
QY 61 ----- 61  
Db 5182 CAACCATAGACTTGTACGTTTGGTTTAACTTTCTAGAACTAATTTAGATTATTATT 5241  
QY 61 ----- 61  
Db 5242 TTGTAGTTTATCTGTCTATGCTATTGTTATTAATCTTTGAATTCAACTGCAATACTTAGA 5301  
QY 61 ----- 61  
Db 5302 TTATCTTGAAGTCTCTCTTTTCTGAGCTGTACAAGCTATGTATGAAATGCCCTACCTCC 5361  
QY 61 ----- 61

Db 5362 AGCATCTTTTAGATTATGTAGGGCCCTTTCTGAGTTTATCAGTTGTATATTGACTGAAGC 5421  
QY 61 ----- 61  
Db 5422 AGCAATGTGTAT 5481  
QY 62 ---ArgGlyTyriValAlaLeuLeuGlnLysLysAspProLysPheCysSerLeuSerArg 80  
Db 5482 ACCAGAGGTATGTTGCTCTCTCTCAGAGAAGGATCCAAATTTCTGCTCTATCTCGG 5541  
QY 81 IlePheHisAspGlnLysLysCysAspGluHisLysAlaSerSerSerProPheSerVal 100  
Db 5542 ATTTTCCATGATCCAGAAAAATATGATGAACACAAAGCTAGTTCAAGCCCAATTTTCTGTA 5601  
QY 101 AlaLysPheArgAspGlnLysLysCysSerLysCysLeuAspLysLeuLysThrSerAspAsn 120  
Db 5602 GCAAGTTTCCAGCATGCGATGCTCGAAGTGTGTAAGTTGAAACATTCAGATAAT 5661  
QY 121 GlyThrAlaProArgThrLeuProAlaLysGlnLysGlnLysThrSerAspGlyCysSerIle 140  
Db 5662 GGAACAGCACCAAGAACTCTTCCCGCAAGCAGATGCGCACAGTGGTGGTCTCCATC 5721  
QY 141 ThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysValSerProSer 160  
Db 5722 ACATTTGTTGGAGCACCTTTTGTGCTCTAGTTGTTGTTCCCAAAAAGTGTCTCTAGC 5781  
QY 161 ThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGlnGlu 180  
Db 5782 ACACAATCATCTCAAGGGAAGAAATGCTGATAGTCAACTCTTCCAAGAGTGTGCAAGAA 5841  
QY 181 GlyAsnAspSerLysCysAsnAlaProSerGlyLysAsnGlyAlaAlaGluAlaAsnThr 200  
Db 5842 GSCAATGACTCTCAAAATGCAATGCGCTTCTGCAAGAAATGGAGCTGCTGAGCCCAATCT 5901  
QY 201 AspSerProMetLys----- 205  
Db 5902 GATTCACCAATGAA-AGGTATGTTAGTAGAGCTTTCAAAATTCCTAAGTAGGATTTT 5960  
QY 206 -----AspLeuGlnGlyProAla 211  
Db 5961 ATTTAAGTATAGATAAATAACTAATGTTGTTGTTGTTTCTCAGATTTCAGAGGCGCAGCC 6020  
QY 212 GlnAsnTyriAspValAlaAlaAsnValSerGluAspAsnThrSerValAspValGlyAla 231  
Db 6021 CAANATTATGATGTGCGACCAATGCTCTGAGGACAACTCTCTGTTGTTGTTGTTGTTG 6080  
QY 232 LeuProGluValProGlnLleThrTrpHisIleGluValAsnGlyAlaAspGlnProPro 251  
Db 6081 TTACCTGAGTTTCCCGCAGATTACATGGCACATAGAAAGTAAATGGTGCAGATCAACTCCA 6140  
QY 252 SerThrProLysLeuSerGluValValLeuLysArgAsnGluAspGluAsnGlyLysThr 271  
Db 6141 TCCACTCCAAAACCTTTCTGAGTGTCTCTCAAAAGAAATGAAGATGAAATCGAAAAACT 6200  
QY 272 GluGluThrLeuValAlaGluGlnCysAsnLeuThrLysAspProLysProValSer 311  
Db 6201 GAAGAGACTCTTTGTTGCTGAGCAGTGCATTTTGACCAAGATCCTTAAGCCCAATGTTCTGA 6260  
QY 292 LysGluAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLysProValSer 311  
Db 6261 AAGGACGTGATCAGTTGCTGAGCGTGCATTTTGACCAAGATCCGAAACCACTGTCT 6320  
QY 312 GlyGlnLysCysGluGlnIleCysAsnGluProCysGluValValLeuLysArgSer 331  
Db 6321 GCGCAGAAATGTGAGCAGATCTGCAATGAGCCATGTGAAGAGTTGTTCTCAAAAAGAGC 6380  
QY 332 SerLysSerLysArgLysThrAspLysLysLeuMetLysLysGlnGlnHisSerLysLys 351  
Db 6381 TCCAAATCTAAGGAGAGCGGATAGAAATTTGATGAGACGACGACGACGACGACGACGAA 6440  
QY 352 ArgThrAlaGlnAlaAspValSerAspAlaLysLeuCysArgArgLysProLysVal 371  
Db 6441 CGCACTGCCAGGCTGATGTTTTCAGATGCAAGCTTTGTCGAGAGAAAGCCAAAAAGGTG 6500

372 ArgLeuLeuSerGluIleIleAsnAlaAsnGlnValGluAspSerArgSerAspGluVal 391  
Db CGGCTTCTATCAGAAATATTAATAGTCTACCGAGTTGAGAGATTCTAGAAGTACGAGATT 6560  
392 HisArgGluAsnAlaAlaAspProCysGluAspAspArgSerThrIleProValProMet 411  
Db CATCGTGAATAAGCCGCTGATCCTGTGAGATGATAGAGTACCATCCCGTCCCGATG 6620  
412 GluValSerMetAspIleProValSerAsnHisThrValGlyGluAspGlyLeuIysSer 431  
Db GAAGTAGCATGATGATATCTGTGTACCAACCATACAGTGGAGAGATGGGTAAATCA 6680  
432 SerLysAsnLysThrLysArgLysFyrSerAspValValAspGlySerSerLeuMet 451  
Db AGTAAGACACAGACAAACGCAATACTCTGATGTTGTAGATGATGATCACTTATG 6740  
452 AsnTrpLeuAsnGlyLysLysLysArgThrGlySerValHisHisThrValAlaHisPro 471  
Db AACTGGCTGAATGGAAGAAAGAAAGAACTGGAAGTGTGCATCACACAGTCTCATCCA 6800  
472 AlaGlyAsnLeuSerAsnLysLysValThrProThrAlaSerThrGlnHisAspAspGlu 491  
Db GCTGGGAATTTGAGCAACCAAAAGTGACACCCACTGGAGTACTCAGCATGATGATGAG 6860  
492 AsnAspThrGluAsnGlyLeuAspThrAsnMetHisLysThrAspValCysGlnHisVal 511  
Db AATGATACTCAAAATGCTCTTGACACAAATATGATCAATAGACAGACAGATGCTGTG 6920  
512 SerGluIleSerThrGlnArgCysSerLysGlyLysThrAlaGlyLeuSerLysGly 531  
Db TCGAAATCTCCACACAGAGGTGCTCATCAAGGGGAAACAGCGGTTTGATGAGGGG 6980  
532 LysThrHisSerAlaAlaSerThrLysFyrGlyGlySerThrArgAsnGlyGlnAsn 551  
Db AAACACATTCAGTCTAGTACCAATATGGTGGTGAAGCACCACAGAAATGGTCAGAAC 7040  
552 IleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrGluAsnSerValLeuSer 571  
Db ATACATGTACTCAGCGCAGAGATCAATGCCAGATGGAACCGAAACTCTGTCTGAGT 7100  
572 HisSerAlaLys----- 575  
Db CACTCGGCAAGGTACGAAATTTGTGAATCATGAGGAATTTTGTCTTTTAAATTGACTG 7160  
575 ----- 575  
7161 AATCAACATTTATCTGTATGAAGGAATAATATTTGGTGCAATAACAATGTTAAGAAATATGC 7220  
575 ----- 575  
7221 ATACATGTTTATATATGCTTTCCAGTCTTCTTCTTCTTACTTATGTTTGTACTCTTT 7280  
575 ----- 575  
7281 TTGTGTGCGTGCGATGTGTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGC 7340  
575 ----- 575  
7341 ACGTGCGTGGCGCAATATCTTTTWTAGACTCATATATAGTATGTTGTAATGGACTGACA 7400  
576 -----ValSerProAlaGluHisAspIleGlnIleMetSerAs 588  
Db TTTTCTCTCTCTCATCTCAGGTTTCTCCAGCTGAGCATGATATCCAAATATGCTGA 7460  
588 pLeuHisGluGlnSerLeuProLysLysLysLysGlnLysLeuGluValThrArgG1 608  
Db CCTTCATGACGAGTCTACCCAGAGAAAGAAAGAGCAAAACTTGAAGTACTCGTGA 7520  
608 uLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysAsnG1 628  
Db AAAAAAGACCAATGATAGATGACATCCCATGGATATTTGTGATGCTAGCTAGCTAANAACCA 7580

628 nHisGluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgIleGlnSerLys 648  
Db GATGAGAGGACGCTTATGACATGACATGACATGATTTCTGACATCAACCGTATTCAATCCAA 7640  
648 eThrThrAlaAspAspCysValIleValAlaAlaLysAspGlySerAspFyrAlaSe 668  
Db GACAACTGCTGATGATGATTTGTAAATAGTAGTCTGCCAAGGATGTTTCAAGATTATGATC 7700  
668 rSerValPheAspThrAsnSerGlnGlnSerLeuAlaSerGlnSerThrGlnLysG1 688  
Db AAGTGTGTTTGACACTAAATCCCAACAAGTCTTGTGGCATCCCAAGTACACAGAGA 7760  
688 uLeuGlnGlyHisLeuAlaLeuThrThrGlnGlnSerProHisProGlnAsnPheGlnSe 708  
Db GTTACAGGCTCATTGTCATTGACCAACAAGAGTCTCCACATCTCCAGAACTTTTCAGTC 7820  
708 rThrGlnGlnGlnGlnThrHisLeuArgMetGluGlnMetValThrIleAlaAlaSerSe 728  
Db TACTCAGAAACAGCAGACACATTTGCGATGGAAGAAATGTCATATTGCTGCAAGCTC 7880  
728 rProLeuPheSerHisHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTrpG1 748  
Db ACCACTATTTTCACTATCATGATGATCATGATATATGCTGAGCACCACCTGAACTTGGG 7940  
748 yArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSerPr 768  
Db CCGTAAGACGCAAGAGAGCTTAACGTGGAGCAATTTAAGGCCACTACAGAAATTTCTCC 8000  
768 oAlaAlaThrCysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerTh 788  
Db AGCAGCACATGTTGGTGTCTCAATTTAGACCTGGTATTCACAGCATGTTACTTGTACTTAC 8060  
788 rHisValMetGlySerSerSerAsnFyrAlaSerArgGlnProValIleAlaProLeuAs 808  
Db TCATGTCTATGGATCTTCCAGCAATTTATGTCATCTGCCCAACAGTAATTTGGCCACTGGA 8120  
808 pArgTyrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAl 828  
Db CCGTATGCTCAAGAGAGCGGTTAAACAGTCCATGCAAGAAATTTTCCAGCAGCAATAGC 8180  
828 aThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValValLeuTyrPr 848  
Db AACCATGGAAGGAGTAAATTTATGTCGTGGAGAAATGCTGCACAAAGTAGTCTTGTAATCC 8240  
848 oLysGluSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSe 868  
Db TAAAGATCCATGCTGCGAGCATCTTCTGAGATGATGATGATCATCATCATTAGCAAG 8300  
868 rPheProAsnTyrGlyThrSerArgAsnGlnMetGluSerGlnLeuHisAsnSerG1 888  
Db CTTTCCCAACTATGGAATCTTAGCAGAACCAAGTGGAGTCTCACTTCAATATTTCTCA 8360  
888 nTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsnG1 908  
Db GTATGCAATATCAGTACAAAGGATCAACAGCAGCATCATATGGCAGTAACCTGAATGG 8420  
908 yLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgPr 928  
Db AAGATTCATTTGACATTCGAGACTTATACGGCATCATGCTGCTGATGATGCTGACAGACC 8480  
928 oLeuArgProHisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIleAlaAs 948  
Db TTTAGCCCAACATCTAGAGTTGGTGTCTTGGCTCTCTGTCGAGAGGAAATTCGAAA 8540  
948 nTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSerThrGlyIleTh 968  
Db CTGTGCGAGAACTGTGGCACAATCTGGTTAAGTTAGAGTGTCTACAGGAAATTAAC 8600  
968 rSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheSerAl 988  
Db ATCCCATCAGATGAACAGAGAGAACATTTTGAAGCCCTGAATTCGGAATTTTTCAGC 8660  
988 aLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAspPheLeuSerAlaAr 1008

8661 AAAATGGAATGCATTGCGTTGGTTCTGTAGTCCAGTGCAGATTTTATCAGCGAG 8720  
1008 gAenSerIleAlaGlnSerTrpThrArgGlyGlySerMetValHisProLeuAspAr 1028  
8721 GAACACATAGCTCAATCTTGGACACAGAGCAGGTAAGGTAAATGTTTCACTCCCTTGATCG 8780  
1028 gPheValArgGlnAspIleCysIleThrAsnIysAsnProAlaAspPheThrIleSe 1048  
8781 GTTTGTGACAGAGGATATCTGTATACTAACAGAACCCAGCTGATTTTACTACAATCAG 8840  
1048 rAsnAspAsnGluTyrMetAspTyrArg 1057  
8841 TAACGATACGAGTATATGATGATTACCGC 8868  
RESULT 4  
LOCUS AP001859 150594 bp DNA linear PLN 27-NOV-2003  
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,  
PAC clone: P0485D09.  
ACCESSION AP001859 BA000010  
VERSION AP001859.1 GI:7630233  
KEYWORDS  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Bakayota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
1  
Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,  
Katayose, Y., Wu, J., Nilmura, Y., Cheng, Z., Nagamura, Y.,  
Antonio, S.A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,  
Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,  
Hamada, M., Harada, C., Hijiishi, S., Honda, M., Ichikawa, Y.,  
Itonuma, A., Iijima, M., Ikeda, M., Ikono, M., Itoh, S., Itoh, T.,  
Itoh, Y., Itoh, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S.,  
Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T.,  
Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaka, H., Nakashima, M.,  
Nakama, Y., Nakamichi, Y., Nakamura, K., Namiki, N., Negishi, M.,  
Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T.,  
Shomura, J., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K.,  
Yamagata, H., Yamane, H., Yoshiki, S., Yoshinara, R., Yukawa, K.,  
Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J.H., Kim, H.I., Eun, M.Y.,  
Yano, M., Jiang, J. and Gojobori, T.  
The genome sequence and structure of rice chromosome 1  
Nature 420 (6913), 312-316 (2002)  
22337376  
12447438  
2 (bases 1 to 150594)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission  
Submitted (19-APR-2000) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail:tsasakienias.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/>,  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
The orientation of the sequence is from 77 to 86 of the PAC clone.  
Genes were predicted from the integrated results of the  
following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as  
SplicePredictor (October 1998 version). The genomic sequence was  
searched against the non-redundant database NRP (PIR, SWISSPROT,  
GENPEPT, PDB) from MAPP DNA bank and the cDNA sequence database at  
RGF. Protein similarities of the coding regions were searched  
against NRP with BLASTP2.0. ESTs represent the identified cDNA  
sequences using BLASTN2.0 with the corresponding DBJ accession no.  
and RGP clone ID.  
This sequence of this clone has an overlap with P0431F01 clone,  
DBJ:AP001550 at the 3' end. This clone ends at the position 57,376  
of P0431F01. Detailed information on overlap and assembly quality  
together with annotation of this entry at  
<http://www.dna.affrc.go.jp:82/genomicdata/GenomeFinished.html>.  
Location/Qualifiers  
1. .150594

/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="1"  
/clone="P0485D09"  
join(7919, .7964, 8092, .8280, 8453, .8473, 8558, .8610)  
/note="unnamed protein product; hypothetical protein"  
/codon\_start=1  
/protein\_id="BAA94767.1"  
/db\_xref="GI:7630234"  
/translation="METRIPIQVRAPBLSVDNRITAFLLSVPGIRDSAFVSPDSIRR  
HRRVLAVLIPVDIRRRVLAIVGFVVVASHMSIVLGCFRIGHVRSPNPQIRLQIKAK  
"  
complement(9235, .9582)  
/note="unnamed protein product; hypothetical protein"  
/codon\_start=1  
/protein\_id="BAA94768.1"  
/db\_xref="GI:7630235"  
/translation="MTARFARVGRERRAPAGRSMEEGGGAHRRRGASPAATAATRIIG  
ARRNFGEGMERGVGYLAIFPGNFIQKRGREGARGCCEMDGEMARVGSARSQSV  
STSAAFVSRFGNF"  
join(10735, .10850, 10933, .10987, 11596, .11653, 12295, .12379,  
12471, .12574, 12958, .13086, 13493, .15302)  
/note="unnamed protein product; Similar to Arabidopsis  
thaliana chromosome 4, BAC clone F4D11; putative  
myb-protein. (AL022537)"  
/codon\_start=1  
/protein\_id="BAA94769.1"  
/db\_xref="GI:7630236"  
/translation="MTSDNGKAPDKGPEPPSPAPGEGEISNEPKRRPLNGRTTGP  
TRRTKGNTPEDDAILSRVQTYNGKWKIAECFPDRTDVQCLHRWQKVLNPELVK  
GPMKDEEIIQVMVKLGPKKWTIAQALPKRICQCRWRWYHNLPMKNEATQEC  
BEITLIHAHRNMGKMAELTKFLPQRTDNSTIKHNWSSVKKKNSVCSQVHDTTVPLG  
PLNYSYNSGSPANTQNSDSCFAVRENSGSCSSVSSALVCHVCHVSSNLDPDQLQED  
CDLQVNFDKNEAHDSDSGMPQACTTAAVASALPVAHVCHVSSNLDPDQLQED  
FAGLNLMDIDEMTVPFADNTVCSNENHRSLEPYDVAMEVPLSMLSDDSGAEQ  
KLFNLSMDTSPNCLSKELMDLSLQGLSDPAVEADISRNHSDSDVYSSADTH  
FLAVYPTQSSSVSMGLADDQSPQSPSPSLICEQAMTDADPDNRPGRKEMFLSQ  
AEVYPTQSSSDAEMFANCPNDRVPSPTMESIPECQDQVTPNEAPEASLEKPE  
SITQVTPADPDQKGLFYEPPEPSLDVPFCDLVATSGDLQFSPGLIGLQKHSMT  
NVTCPMLWSFTDSTGTLKSAASFCITPSILKRRHDLSPIDPKIEKKGIT  
EKDGMVSTSGTGTSCINAKDQALITVIRIERSSASSSKLKLVSDEKENLG  
YTFETQDQSGAGNDEHDEOTTGERSATNVAINDLSGNLVSTSS"  
complement(join(18608, .18823, 19905, .20031, 20072, .20220,  
20774, .21113, 21407, .21738, 21975, .22508))  
/note="unnamed protein product; ESTs  
AU082761(S5084), D42006(S5084) correspond to a region of  
the predicted gene.  
Similar to caenorhabditis elegans chromosome 3, coemid  
C17G10; alcohol dehydrogenase/ribitol dehydrogenase.  
(U28739)"  
/codon\_start=1  
/protein\_id="BAA94770.1"  
/db\_xref="GI:7630237"  
/translation="MAAARAGGGGAGVDPSPVPSVKTFFVAHYRHRVDRKNVYE  
IHQVEGQFQLSDRLFDAPWPAABAVSPVCDGHVFLLYRLVYRHAARASSSS  
SSSAPLTAGRAESWANTCDLFSVLHGVVNMQLPQWLMDWDEYVYQFQFCQYRA  
KLKNSDDEHQKQFQKANNVYGVNLQALVSKMIQALEREKELEQFTATDGY  
DYQGSNVLMGLYSGIMGLIRHCLIGDYTGKLCAPIDLNQOQVTVIIGSHISA  
IYHGFANLMMRYAERLREFNKILLYLKTQHKSPQYDQILKKNQMYAFPLAVC  
LDQPLTWNQDQAVLQKLFLYEVKQQLSGIRSLKLYSTITIAKLQMYEVEAT  
LRMKNLSQYLRSILMYTKHMAVONNGKIVSSADPDFYIKCISIRTPKLTSKN  
LAHLHQGVKTVRKPSPLSFSFSPFGGICNGSYENGIYKYNRNSNEPQIR"  
join(23045, .23145, 24592, .24670, 25118, .25216, 25302, .25415,  
25608, .25750, 26364, .26445, 27073, .27158, 27527, .27660,  
29332, .29917)  
/note="unnamed protein product; Similar to Arabidopsis  
thaliana chromosome 2 BAC T19C31; unknown protein.  
(AC004583)"  
/codon\_start=1  
/protein\_id="BAA94771.1"

```
/db xref="GI:7630238"
/translation="MAGSMYDRATSDMLIGPQWAKNNEICIDNRPQSGQKDVVKALK
LTHGNPKVQAILLLETAIKNGCDIFHMVHAERDVLHEWVKIYKKSGDNVREKV
KRWIDTPKQAGPRARYQVYAAHDLVRAAGAPKRDPAPIFNGSGRGNWES
PQQDARSAGCNDPFAIMSEIQNARGMDVLAEMLNLDLFGNREGLRQEVIVELVD
QRTYKQVLLVNDABELMSQGLALNDLQRLAKDAAAGLAVRVRKPKSLQ
ALVDDSNQDQSKQALVDIEDPTQBTNKESQASVQSFEGALPAPVPSNGS
ATAPKSDGLGDLWDDNPSITENSALALVPDVPDVSQNALAVDIIFSQNNTT
NLIAPKDPGVNNSSTLPGSQYNAPTQHPLOQQPQOQVGLPNGRAVNPCTSYPT
SSGNGOIANNAATPPAQCTVNYGIAVLVYIISQPSYTIIPIGQNHACPLCSV"
30166..30675
/notes="unnamed protein product; ESTs
AU031545(I561825),AU031546(I561825) correspond to a region
of the predicted gene.
hypothetical protein"
/codon_start=1
/protein_id="BAA94772.1"
/db xref="GI:7630238"
/translation="WYNQPGAMQSPQAVTQXQGFANQFSGMPPHSMQPPQPM
QPSMPGAQVQMYMMPGQFPAFMQPMQYRQVRLVQQQAQYISNGSR
PTTYAGMDLDSQMYGLSMQDSSYNGMNSPSTTTPSSSSMQSPKPKDLGLDLS
IAKTQNKXA"
complement(join(34629..34795,36831..36867,38803..38895,
41780..41830))
/notes="unnamed protein product; hypothetical protein"
/codon_start=1
/protein_id="BAA94773.1"
/db xref="GI:7630240"
/translation="MTVYRCTTQRRLQEGRIQCDIRGKILSGKGFPSQNRKNTGI
EIEVGLIDACVIVTVKEPENERRIKKHGRLGLSLCDRLGPKLVDSPTKVVTPY
VQPRGHTLPHRG"
join(46367..46423,47537..47769,48031..48153,50082..50299,
51762..51866,52091..52204,52433..52532,52784..53004,
54084..54516,54601..55709,56020..57468)
/notes="unnamed protein product; hypothetical protein"
/codon_start=1
/protein_id="BAA94774.1"
/db xref="GI:7630241"
/translation="MNGELIKREBELLEKLIHATFPHPPOIPVGAATAAATAIA
WICAPVAANAICAPSPVDGSPCLRAVPPPNRQTRPSVAPSPIDRHIPL
IGTMEKDLIFMTHGVRVADDEVRSRARTELKIRGVGLMDVCGYPIHVVGDG
FMCVRAVHRMNGSGRLALLTCLTKANCLLAKMLFLFKKQTFELGNLBEAASG
ESTARVDRAPPLRREGCSKRPKSPCAPSLGLELLASMGTKALIFHVAEBKIPSF
AGILAAFWNLFLAIILPARISGALIIFHELIIVDIAACLVDSLLVQVDEIVAV
DOGARVVGTCMLARGTCAVAPLELTATPRQDAAEAGVDBPAQOCHEPSIRGY
VALLOKDKFCSLRIIPHQKQKDEHKASSPFSVAKFRWDCSKLKLKTSIDNGT
APRTLPKQGTSDGCSITFVRSITFVPSVSGSKVSPSTOSSQGNADSTLPKSVQ
GNDSKNAPSGKNGAAEANTDSPMKDLQGAQNTDVAANVSEDNISVDVGLAFVFOI
TWIEVNGADQPTSTPKLSEVLKRNEDENGKTEETLVABQCNLTDPNPMGSKERDQ
VAEQCNLTDPKPVSGQKCBQICNEPCBEVVLARSKSKRTDKLAKKQSHSKRTA
QADVSDAKLCRRPKPKVRLSEIINANOVEDSRSDVHRENAADPCDDRSITIPVME
VSMIDIPVSNHTVGEDGLKSSKNKTKRYSDVDDGSLMNLNGKKRTGSHVHTVAH
PAGNLSNKVTPASTQHDNDENTENGLDTNMEKTDVCOHVSEISTQRCSSKGTAGL
SKGTHSASTKYGESSTRNGQNIHLVSAEQCOMETENSVLSHSAKVSFAEDIQIM
SDLHEQSLPKKKKQKLEVTREKQTMIDDIIPMDIIVELLAKNQHQRQMTETDCSDLR
IQSKTTADDDCVIVAAXGSDYASSVEDTNSQKSLASQSTKELOGLHALTTQESPH
PONFQSQEQTHLREEMVTIAASSPLFSHDDQYIAEAPTEHWGRKDAKCLTWQF
KATTRNSPAATCGAOPRPGVQAVDLSTHVWSSSNVSPQVIAPLDRVAERAVNQV
HARNPSTIAEASKLCDRNRAGVVLVPEKSMPTALLRMDPSTLASFVNYGTSS
RNQWESOLNSQVARNKQSTSTSYGSLNKGILPTEFDLSRQHLDLHRLRPHR
VGLVGLQLEKIANWNSCNGTQSGIKLGVTGITSQHMNRKEHFEALNSGFSAKWNA
LQGSVSSADPUSARNSIAQSWTRGKRVHPLDRFVRQDICIITNKNPADFTIISND
NEYMDYR"
complement(join(58634..58667,58796..58902,59018..59083,
59368..59432,59585..59648,60333..60442,60620..60707,
60835..60945))
/notes="unnamed protein product; ESTs
AU082698(R0665),D23963(R0665) correspond to a region of
the predicted gene.
Similar to Arabidopsis thaliana rac GTP binding protein
Arac7. (AF079484)"
```

Alignment Scores:

	Pred. No.:	3,34e-301	Length:	150594
Score:		5191.00	Matches:	1055
Percent Similarity:		68.06%	Conservative:	0
Best Local Similarity:		68.06%	Mismatches:	2
Query Match:		93.94%	Indels:	495
DB:		8	Gaps:	3
US-09-721-114-2 (1-1057) x AP001859 (1-150594)				
QY	1	MetGluLeValaValaAspGlnGluGlyAlaAa:gvValValGlyThrAsnCysMetLeu	20	
DB	52820	ATGGAGATTGTTGCAGTAGATCAGGAGGAGCTCGTGTGTTGGCAGCAACTGTATGCTT	52879	
QY	21	AlaArgGlyGlyThrGlyAlaValaValaProValLeuGluLeuThraThrProArgGln	40	
DB	52880	GCTCGTGTGGAACTGGTCTGTAGCGCCAGTGTGTGGAGCTGACAGCGCGCTGTCAG	52939	
QY	41	AspAlaAlaAlaGluAlaGlyValaAspGluProAlaGlnHisGlnCysGluHisPheSer	60	
DB	52940	GATGCGCCGCTGAAGCTGTGTAGAGAACCGGCACACCAATGCGAGCATTTCTCC	52999	
QY	61	Ile-----	61	
DB	53000	AT-AAGGTAATCATTTTCTGTATTTCCTCAATCCAGTATCGCGTGTGGATGAATAAGAA	53058	
QY	61	-----	61	
DB	53059	TGGCATGTATGCCCATATTGTCATGTTTGTATGGAAGATGATGATGATCGTGGTTTTG	53118	
QY	61	-----	61	
DB	53119	CACAGTTTGTGTGGGACTTATATGGTCATCTGTGTTTGTAGGATCGTATACACTGGGTC	53178	
QY	61	-----	61	
DB	53179	GACATGCTTATGACTTTTGGTGGATTAGGAAGTCAATACACTACTAGCTCTATAT	53238	
QY	61	-----	61	
DB	53239	CTAGCCATGTGAACATCATTTATGCCATAGACAGCTAGCAGGCTAGCAGCAAAAATATA	53298	
QY	61	-----	61	
DB	53299	TATAATATTGTCATATATGTTGGTGTTCATGATATCTTATATCTACTACGTACATCCATTA	53358	
QY	61	-----	61	
DB	53359	ATATCTTCAATGTATGAATCTGAGCAGCATGATTTGTGAGTGTCTACATATGCAATGCTGT	53418	
QY	61	-----	61	
DB	53419	ATGTGTTCATTAGGTGTTTGCATCATATTTGTTGTTGGGGTGGCGATGCAATTTAT	53478	
QY	61	-----	61	
DB	53479	CAGGCCATGCTGTAGGCTGTAGTAGATATTTGTGTTGTATATTTCTGTGTGAACAA	53538	
QY	61	-----	61	
DB	53539	GCTGATTACTAATGAATAATTAACCTTTTGGGGTACACTCATATATTGGGCCCTACATTTT	53598	
QY	61	-----	61	
DB	53599	TGTAAATCAATTTTCTTCTGCTGCTAGGTTTCAGCATAAACTTTTATCATTAAGCATGTT	53658	
QY	61	-----	61	
DB	53659	TACATCTAGGAGATTTCTTAGAAGTGTGTTCTTCATATTTTGCATTATTTGTTGTTG	53718	
QY	61	-----	61	
DB	53719	ATAGTCCATTTATTTTAAAGCCTTTTCAATTTGTTAGAGATTTCTAGAGATGATATAT	53778	



55938	Db	ACGTGCGTGGCGCAATATTCTTTTATTAGACTCATATTAATAGTGAATTGTAATAGGACTGACA	55997
576	Qy	-----ValSerProAlaGluHisAspIleGlnIleMetSerAs	588
55998	Db	TTTTCCCTCAATTTCTCATCTCAGGTTTCTCCAGCTGAGCATGATATCCAAATATATGCTCTGA	56057
588	Qy	pLeuHisGlnSerLeuProLysLysLysLysLysLysLysLeuGluValThrArgG	608
56058	Db	CCTTTCATGAGCAGAGTCTACCCAAAGAAGAAAGACGAAAAACCTTGGAAGTGACTCGTGA	56117
608	Qy	uLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuLeuAlaLysAsnG	628
56118	Db	AAACACAGCCATCATAGATGACATCCCATGGATATTGTTGAACCTGTAGCTTAAACCA	56177
628	Qy	nHisGluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgIleGlnSerLys	648
56178	Db	GCATGAGAGCAGCTTATGACTCAGAGCTGATTTGTTCTGACATCAACCGTATTCAATCCAA	56237
648	Qy	sThrThrAlaAspAspCysValIleValaIalAlaLysAspGlySerAspTyrAlaSe	668
56238	Db	GACAACTGCTGATGATGATTTGTTAATAGTAGCTGCCAGGATGGTTAGATTATGCATC	56297
668	Qy	rSerValPheAspThrAsnSerGlnGlnLysSerLeuAlaSerGlnSerThrGlnLysG	688
56298	Db	AAGTGTGTTTGACACTAAATTTCCCAACAGAAGTCCTTGGCATCCCAAGAGTAGACAGAAGCA	56357
688	Qy	uLeuGlnGlyHisLeuAlaLeuThrThrGlnLysSerProHisProGlnAsnPheGlnSe	708
56358	Db	GTTACAGGGTCAATTTGGCATTTGACCAACAAGAGTCTCCACATCTCCAGAACTTTTCAGTC	56417
708	Qy	rThrGlnGluGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerSe	728
56418	Db	TACTCAGGAACACAGACACATTTGCGATGGAGAAATGGTTCATATTGTGCGAAAGCTC	56477
728	Qy	rProLeuPheSerHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTyrG	748
56478	Db	ACCACATATTTCATCATCATGATGATCAGTATATTGCTGAAGCACCAACTGGAACATTTGGG	56537
748	Qy	YArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSerPr	768
56538	Db	CCGTAGGAGCGCAAGAGCTTAACGTGGAGCAATTAAGGCCACTACAGAAATTTCTCC	56597
768	Qy	oAlaAlaThrCysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerTh	788
56598	Db	AGCAGCAACATGTGGTGTCTCAATTTAGACTCGGTATCCAAAGCAGTTGACTTGACTCTAC	56657
788	Qy	rHisValMetGlySerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAs	808
56658	Db	TCATGTCATGGGATCTTCAGCAATATATGCATCTCGCAACCGTAATTGCGCACATGGA	56717
808	Qy	pArgTyrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAl	828
56718	Db	CCGCTATCTGAAAGAGCGGTTAAACAGGTCCATGCAAGAAATTTTCCAGCAACAATAGC	56777
828	Qy	aThrMetGluAlaSerLysLeuCysAspArgAsnAlaGlyGlnValValLeuTyrPr	848
56778	Db	AAACATGGAAGCAGATGATGTTATGTGATCGGAGAAATGCTGGACAAGTAGTCTTGTATPCC	56837
848	Qy	oLysGluSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSe	868
56838	Db	TAAAGATTCATGCTCTCGCGGCATCTTCTGAGATGATGGATCCATCAACATTAGCAG	56897
868	Qy	rPheProAsnTyrGlyThrSerSerArgAsnGlnValGluSerGlnLeuHisAsnSerG	888
56898	Db	CTTTCCCAACATAATGGAACCTTTAGCAGGAACAGATGGAGTCTCAACTTCTAATAATCTCA	56957
888	Qy	nTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsnG	908
56958	Db	GTAATGCACATATCATGATCAATAGGATCAACACAGCACTCATATGCGATGATACCTGAAATGG	57017
908	Qy	YLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgPr	928
57018	Db	AAAGATTCATTCATTCATTCGAGACTTATCAAGGATCAGCTGCTGATGATCTCCACAGACC	57077

Qy	928	OLeuArgProHisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIleAlaAs 948
Db	57078	TTTACGCCCCACATCTTACAGAGTGCGTGTGGCTTCCTTGCACAGAAGAAATTGCAAA 57137
Qy	948	nTPSerGluAsnCysGlyThrGlnSerClyTrpLysLeuGlyValSerThrGlyIleTh 968
Db	57138	CTGGTCGGAGACTGTGGCACCAATCTGGTTATTAAGTTAGAGTGTCAACAGGNATAAC 57197
Qy	968	rSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheSerAl 988
Db	57198	ATCGCATTCAGATGAACAGAAAGACAATTTTGAACCCCTGAAATTCCTGGAATGTTTTCAGC 57257
Qy	988	aLYSTpAsnAlaLeuGlnLysGlySerValSerSerSerLalaspPheLeuSerAlaAr 1008
Db	57258	AAATGGAATGCATTCAGATTGGTCTTCTGTAGCTCCAGTCAGAGATTTTTTATCAGCGAG 57317
Qy	1008	gAsnSerIleAlaGlnSerTrpThrArgLysGlyLysMetValHisProLeuAspAr 1028
Db	57318	GAAACAGCATAGCTCAATCTTGACACAGAGGCAAGGGTAAAAATGGTTTCATCCCTTCGATCG 57377
Qy	1028	gPheValArgGlnAspIleCyValleThrAsnLysAsnProHlaspPheThrIleSe 1048
Db	57378	GTTTGTGAGACAGGATATCTGTATTAATCAACAGAACCCAGCTGATTTTATCAATCAG 57437
Qy	1048	rAsnAspAsnGluTyrMetAspTryArg 1057
Db	57438	TAAAGTAACGAGTATATCGAATTCGCCG 57465
RESULT 5		
LOCUS	AKI03364	
DEFINITION	Oryza sativa (japonica cultivar-group) cDNA clone J03312EN23, full insert sequence.	linear PLN 24-JUL-2003
ACCESSION	AKI03364	
VERSION	AKI03364.1 GI:32988573	
KEYWORDS	PLI CDNA; CAP trapper.	
SOURCE	Oryza sativa (japonica cultivar-group)	
ORGANISM	Oryza sativa (japonica cultivar-group) Rukayorca; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.	
REFERENCE	1	The Rice Full-Length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team, Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., Hotta I., Koizumi K., Namiki T., Ohneda E., Yahagi W., Suzuki K., Li C., Ohtsuki K., Shishiki T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group, Otsuo Y., Murakami K., Tada Y., Sugano S., Fujimura T., Suzuki Y., Tsunoda Y., Kurosaki T., Kodama T., Masuda H., Kobayashi M., Xie Q., Lu M., Narikawa R., Sugiyama A., Mizuno K., Yokomizo S., Niikura J., Ikeda R., Ishibiki J., Kawamata M., Yoshimura A., Miura J., Kusumegi T., Oka M., Ryu R., Ueda M., Matsubara K., RIKEN, Hara J., Carninci P., Adachi J., Aizawa K., Arakawa T., Fukuda S., Kawai A., Hashidume W., Hayatsu N., Imotani K., Ishii Y., Itoh M., Kagawa I., Kondo S., Konno H., Miyazaki A., Osato N., Ota Y., Saito R., Sasaki D., Sato K., Shibata K., Shinagawa A., Shiraki T., Yoshino M. and Hayashizaki Y. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
JOURNAL MEDLINE	Science 301 (5631), 376-379 (2003)	
PUBMED	22752273	
REFERENCE	12869764	
AUTHORS	2 (bases 1 to 3626)	
	Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Doi K., Fujisawa T., Fukuda S., Hanagaki T., Hara A., Hashizume W., Hayashida K., Hayashizaki Y., Hayatsu N., Hiramoto K., Hiroaka T., Horii F., Hotta I., Iida J., Iida Y., Ikeda R., Imamura K., Imotani K., Ishibiki J., Ishii Y., Ishikawa M., Itomura M., Kagawa I., Kagawa S., Katoh H., Kawagashira N., Kawai J., Kawamata M., Kikuchi S., Kishikawa-Hirose T., Kishimoto N., Kobayashi M.,	





QY 401 GluAspAspArgSerThrIleProValProMetGluValSerMetAspIleProValSer 420  
 DB 1847 GAGGATGATAGAGTACCATCCCGTCCCGATGGAGTAGCATGATATCTCTGTAGC 1906  
 QY 421 AsnHisThrValGlyGluAspGlyLeuLysSerSerLysAsnLysThrLysArgLysPyr 440  
 DB 1907 AACCATACAGTGGGAGAGATGGTTAAATCAAGTAAGAACCAAGCAAAACCAATAC 1966  
 QY 441 SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysValAsp 460  
 DB 1967 TCGATGTTGTAGATGATGATCATCTTATGACTGGCTGATGGAATGGAATAAGAAAGA 2026  
 QY 461 ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
 DB 2027 ACTGGAAGTGTGCATCACACAGTTGCTCATCCAGCTGGGAATTTGAGCAACAAAAGTG 2086  
 QY 481 ThrProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAspThr 500  
 DB 2087 ACACCCACTCGGAGTACTCGACATGATGATGAGAAATGATGAAATGCTTGTGACACA 2146  
 QY 501 AsnMetHisLysThrAspValCysGlnHisValSerGluLysSerThrGlnArgCysSer 520  
 DB 2147 AATATGANTAGACAGATGCTGTGCAGCATGTATCAGAAATCTCCACAGAGGTGCTCA 2206  
 QY 521 SerLysGlyLysThrAlaGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLys 540  
 DB 2207 TCAAGGGGAAAACAGCGGTTTGAGTAAGGGGAAAACACATTCAGCTGTGTAGTACCAA 2266  
 QY 541 TyrGlyGlyGluSerThrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGln 560  
 DB 2267 TATGGTGGTGAAGCACAGAAATGTGCAGACATATGATCTACTCAGCGCAGAGATCAA 2326  
 QY 561 CysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLysValSerProAlaGlu 580  
 DB 2327 TGCAGATGAAACCGAAAACCTGTCTCAGTCACTCGGCAAAAGTGTTCCTCCAGCTGAG 2386  
 QY 581 HisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysLys 600  
 DB 2387 CATGATATCAAAATATGTCTGACCTTCATGAGCGAGTCTACCCAGAAAGAAAGAGAG 2446  
 QY 601 GlnLysLeuGluValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIle 620  
 DB 2447 CAAAAACTTGAAGTGACTCGTGGAAAAACAGACCATGATGATGATGATGATGATGAT 2506  
 QY 621 ValGluLeuLeuAlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSer 640  
 DB 2507 GTTGAACTGTAGCTAAACCCAGCATGAGAGGCGAGCTTATGACTGAGACTGATTTCT 2566  
 QY 641 AspIleAsnArgIleGlnSerLysThrThrAlaAspAspAspCysValIleValAlaAla 660  
 DB 2567 GACATCAACCGTATTCATCCAGACCACTGCTGATGATGATGATGATGATGATGATG 2626  
 QY 661 LysAspGlySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGlnLysSerLeu 680  
 DB 2627 AAGGATGGTTTCAGATTATGATCATCAAGTGTGTTGACACATTAATCCCAACAGAGTCT 2686  
 QY 681 AlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGluSer 700  
 DB 2687 GCATCCCAAGTACAGAGAGAGTATACGGGTCTATTGGCATTTGACCAACAGAGTCT 2746  
 QY 701 ProHisProGlnAsnPheGlnSerThrGlnGlnGlnGlnThrHisLeuArgMetGluGlu 720  
 DB 2747 CCACATCTCAGAACTTTTCAGTCTACTCAGGAACAGCAGACACATTTGCGGATGGAAGA 2806  
 QY 721 MetValThrIleAlaAlaSerSerProLeuPheSerHisHisAspAspGlnTyrIleAla 740  
 DB 2807 ATGTCTACTATTGTCTGCAAGCTCACCATTATTTTTCATCATCATGATGATCATGATG 2866  
 QY 741 GluAlaProThrGluHisThrGlyArgLysAspAlaLysLysLeuThrTrpGluGlnPhe 760  
 DB 2867 GAAGCACCACTGACATTTGGGCGCTGAGGCGGCAAGAGCTAACCTGGGACCAATTT 2926

QY 761 LysAlaThrThrArgAsnSerProAlaAlaThrCysGlyAlaGlnPheArgProGlyIle 780  
 DB 2927 AAGGCCACTTACAGAAATTTCTCCAGCAGCAACATATGGTGTCTCAATTTAGACTGTATC 2986  
 QY 781 GlnAlaValAspLeuThrSerThrHisValMetGlySerSerSerAsnTyrAlaSerArg 800  
 DB 2987 CAAGCAGTTGACTTCACTTCTACTCATGTATGGGATCTTCAGCAATTTATGCACTCCG 3046  
 QY 801 GlnProValIleAlaProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAla 820  
 DB 3047 CAACCAAGTAATTTGGCCCACTGACCGCTATGTGAAAGAGAGCGTTTAAACCAAGTCCATGCA 3106  
 QY 821 ArgAsnPheProSerThrIleAlaThrMetGluAlaSerLysLysCysAspArgAsn 840  
 DB 3107 AGAAATTTTCCCAAGCACATAGCAACCACTGGAAGCGAGTAAGTTATGTATGATCGGAAAT 3166  
 QY 841 AlaGlyGlnValValLeuTyrProLysGluSerMetProAlaThrHisLeuLeuArgMet 860  
 DB 3167 GCTGGACAGTAGTCTTGTATCTTAAAGATTCATGCTGCGACGCACTCTCTGGAATG 3226  
 QY 861 MetAspProSerThrLeuAlaSerPheProAsnTyrGlyThrSerSerArgAsnGlnMet 880  
 DB 3227 ATGGATCCATCAACATTAGCAAGCTTCCCAACTATGGAACCTTCTAGCAGGAAACCAATG 3286  
 QY 881 GluSerGlnLeuHisAsnSerGlnTyrAlaHisAsnGlnTyrLysGlySerThrSerThr 900  
 DB 3287 GAGTCTCAACTTCATTAATTTCTCAGTATGCAATATATCATGACAAAGGATCAACAGCACA 3346  
 QY 901 SerTyrGlySerAsnLeuAsnGlyLysIleProLeuThrPheGluAspLeuSerArgHis 920  
 DB 3347 TCATATGGCAGTAACCTGAATGGAAGATTCATTCATTCATTCATTCATTCATTCATTCAT 3406  
 QY 921 GlnLeuHisAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlySer 940  
 DB 3407 CAGCTGCATGATCTGCACAGACCTTTTACGCCACATCTCTAGAGTTGGTGTGTGCTGCC 3466  
 QY 941 LeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLys 960  
 DB 3467 TTGCTGCACAGAGAAATGCAAACTGGTGGAGAACTGTGGCACAACAATCTGCTTATAAG 3526  
 QY 961 LeuGlyValSerThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAla 980  
 DB 3527 TTAGGAGTGTCAACAGGAATAACATCGCATCAGATGAAACAGAAAGGAACATTTTGAAGCC 3586  
 QY 981 LeuAsnSerGlyMetPheSerAlaLysTrpAsnAlaLeu 993  
 DB 3587 CTGAATCTTGGATGTTTTCAGCAAAATGGAATGCATTG 3625

RESULT 6  
 AX590041  
 LOCUS 3291 bp DNA linear PAT 24-JAN-2003  
 DEFINITION Sequence 223 from Patent WO02081695.  
 ACCESSION AX590041  
 VERSION AX590041.1 GI:27901181  
 KEYWORDS Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 SOURCE Arabidopsis thaliana  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 REFERENCE 1.  
 AUTHORS Zhu, T., Glazov, E.A., Meins, F., Wang, X. and Chang, H.S.  
 TITLE Genes that are modulated by posttranscriptional gene silencing  
 JOURNAL Patent: WO 02081695-A 223 17-OCT-2002;  
 Syngenta Participations AG (CH)  
 FEATURES  
 Location/Qualifiers  
 1..3291  
 /organism="Arabidopsis thaliana"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:3702"

ORIGIN  
 Alignment Scores:



Pred. No.: 2,32e-14 Length: 3291  
Score: 393.00 Matches: 248  
Percent Similarity: 36.83% Conservative: 191  
Best Local Similarity: 20.81% Mismatches: 401  
Query Match: 7.11% Indels: 352  
DB: 6 Gaps: 58

US-09-721-114-2 (1-1057) x AX590041 (1-3291)

QY 55 GlnCysGluHisPheSerIleArgGlyThrValAlaLeuLeuGlnLysAspProLys 74  
Db 73 AAATGTCATCTTTCCATGGTGGATTCGTAGCTGAACTCGTGAGAGAGACCTTAGA 132  
QY 75 PheCys-----SerLeuSerArgIlePheHisAspGlnLysCys 88  
Db 133 AAATGTCGCGCTTTTCAGAGAGAGTGTAGTTAGTA----- 171  
QY 89 AspGluHisLysAlaSerProPheSerValAlaLysPheArgArgIlePheAspCys 108  
Db 172 GACCAACAAGCTATCTCTCTACTTATCTGTTCCTGCTGAACTTGTAGTGGTGGATGTT 231  
QY 109 SerLysCysLeuAspLysLysThrSerAspAsnGlyThrAlaProArgThrLeuPro 128  
Db 232 ATGAGCTGCATCAAGATATA-----GATGCTCATGGGCCAAAGATTGTGGA 279  
QY 129 AlaLysGluAsnGlyThrSerAspGly---CysSerIleThrPheValArgSerThrPhe 147  
Db 280 CTCATTCAAACTCAAAGCTATGGAAGTCTCTCTCTGTTATCGAAGTAAAGCAAGTTC 339  
QY 148 ValProAlaSerValGlySerGlnLysValSerProSerThr----- 161  
Db 340 AATTCGCTAATCATTCATGACGAGAGAAAGAAAGAACTGATATTGCAGATAATGCT 399  
QY 162 ---GlnSerSerGlnGlyLysAsnAlaAspArg-----SerThrLeuPro 175  
Db 400 ATTGAGAGAAAGTGGGTGAAACTGTGAGATGATGATGATGATGATGATGATGATGATG 459  
QY 176 LysSerValGlnGluGlyAsnAspSerLysCysAsnAlaProSerGlyLysAsnGlyAla 195  
Db 460 AAGAAGCAGCT-----GGTCGACCTATGGT----- 486  
QY 196 AlaGluAlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnThrAsp 215  
Db 487 ---GCTTCTAATGTAGGCAAGAGAGAGAAAGCTGTGAGTCCGAGCAG----- 534  
QY 216 ValAlaAlaAsnValSerGluAsp-----AsnThrSerValAspValGlyAlaLeu 232  
Db 535 GTAGGAAACAACAGATCTAAGAAAGAAAGAACTAAACCAACCTGATGATATTAGCAGCTG 594  
QY 233 ProGluValProGlnIleThrPheIleGluValAsnGlyAla----- 247  
Db 595 AAAGAGAAACAAATGTGGATCAGGCTGTGACAAAGCTTCGCTCATCTGAAATTCGCTGT 654  
QY 248 -----AspGlnProSerThrProLysLeuSerGluValValLeuLysArgAsn 264  
Db 655 GTGTTGAGGATACACCACTTAGGCAACCAAGATCATTAAGGCATTCCGGTCTGATG 714  
QY 265 GluAspGluAsnGlyLysThrGluThrLeuValAlaGluGlnCysAsnLeuThrLys 284  
Db 715 GAATGCGATACCGGTCATCAGAAAGTATAAATCTGCT----- 753  
QY 285 AspProAsnProMetSerGlyLysGluArg-----AspGlnValAlaGluGlnCys 301  
Db 754 -----ATGATGGTGTTCAGCGTAGGAATCTCGCAAGGTTCTCTACTCAGT 801  
QY 302 AsnLeuThrLysAspProLysProValSerGlyGlnLysCysGluGlnIleCysAsnGlu 321  
Db 802 GAGTTGCTGGTATACAAAGAACCCAGGTGGTAGTAACATCAGNAAA----- 849  
QY 322 ProCysGluGluValValLysArgSerSerLysSerLysArgLysThrAspLysLys 341  
Db 850 -----GAAGAGTCTGCTTTGAGAGAGGAATCAGTTAGAGGTGAAAG-----AGAAAG 897

QY 342 LeuMetLysLysGlnGlnHis---SerLysLysArgThrAlaGlnAlaAspValSerAsp 360  
Db 898 TTGTTACTGAAACAATAATTATGTCAGCGCGGATATTGAGTACAATGGGTGCAACCTCTGAA 957  
QY 361 -----AlaLysLeuCys----- 364  
Db 958 AATGCTTCCAAAAGTTGTGACTGTGATCAAGGTAAATAGTCAATCACTGATAGTGGGTT 1017  
QY 365 -----ArgArgLysProLysLysValArgLeuLeuSerGluLeu 377  
Db 1018 GACAGAACTCCATTTAAAGGTAAGCAGAGAAACAGAGATTTCAGGTTGTTGACGAGTTT 1077  
QY 378 IleAsnAlaAsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAla 397  
Db 1078 GTACCATCCTCTCTGTGAAATCTCAAA-----GAAGGTATCAAGGACATGATGCA 1131  
QY 398 AspProCysGluAspAspArgSerThr----- 406  
Db 1132 GATCTCTAGTAG-----AGATCAACTCTCTCGGCACCTTTTATTCACCTGGAACAGTTCT 1185  
QY 407 IleProValProMetGluValSerMetAspIleProValSerAsnHisThrValGlyGlu 426  
Db 1186 GTTCTCTGCTCT-----COGGGTACTCAGAGAAACA-----GAG 1218  
QY 427 AspGlyLeuLysSerSerLysAsnLysThrLysArgLysThrSerAspValValAspAsp 446  
Db 1219 AGGAAGCTCAGTTTACCAGAGAGAGACAAAGAG-----CCTGTATCGATAAT 1269  
QY 447 GlySerSerLeuMet-----AsnTrpLeuAsnGlyLysLys-----LysArg 460  
Db 1270 GGAAGAGCACTGTGATCAGTTTATGATACCGCATTTGATGGAAGTCAAGTTAACTCGCAT 1329  
QY 461 ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
Db 1330 ACTGCTCTCTCCATGAAACACACAGATATCCAAACTCGAGACTTATTGATGGGAAAGGGTG 1389  
QY 481 ThrProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAspThr 500  
Db 1390 -----GGCGGTTATTTCAC 1404  
QY 501 AsnMetHisLysThrAsp-----ValCysGlnHisValSerGluIleSerThrGlnArg 518  
Db 1405 AACCGTTTGGCTTCAGATGGATATTTCAGAAATATCTCTCTCAGGTTAATGATAAGCGG 1464  
QY 519 CysSerSer-----LysGlyLysThrAlaGlyLeuSer 529  
Db 1465 ATAACATCTTGCCATTTGCCAAGCAATGATTATGAGGTCAAGAGACCGGAAACCAAC 1524  
QY 530 LysGlyLysThrHisSerAlaAlaSerThrLysThrGlyGlyGluSerThrArgAsnGly 549  
Db 1525 TGTCTTCAGATTTTGTCTCTCTTAAATCCAGCTCAGGTGGATGTTGAGAACTGGA 1584  
QY 550 GlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrGluAsnSerVal 569  
Db 1585 -----GTAGATATTGTTGACTTCAGAAACCAACCAACCAATACAAACAGATCGTCTTC 1638  
QY 570 LeuSerHisSerAlaLysValSerProAlaGluHisAspIleGlnIleMetSerAspLeu 589  
Db 1639 TCGAACTTGAAGCTTAAGATACCCCTTCTTCTACTGAGTTGCGGATTATCTCGGTG 1698  
QY 590 -----HisGlu 591  
Db 1699 CTGCAAAAAGAGTCTTCTGTCAGATAGAAAGGGAAGACTTTATGTTCCAGAACAT 1758  
QY 592 GlnSerLeuProLysLysLysGlnLysLeuGluValThrArgGluLysGlnThr 611  
Db 1759 CATGAGACCAAGAGCCAAAGTCAGATAGAAAGGAGACTACCACTGAAGAGCAAAAC 1818  
QY 612 MetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysAsnGlnHisGluArg 631  
Db 1819 AAC---GATGATATTCCAAATGGAGATAGTGGAGCTCATGGCCCAAAACCAAGTACGAGAG 1875  
QY 632 GlnLeuMet---ThrGluThrAspCysSerAspIleAsnArgIleGlnSerLysThrThr 650

1876	DB	1876	FTGCTTCCCGACAAAAGAGAGATGTTAGCAACAAACAGCCATCAACAAGAAAACAGACAC	1935
651	QY	651	AlaAspAspCysValIleValAlaAlaLysAspGlySerAspTyrAlaSerSerVal	670
1936	DB	1936	AAATCCAGAAATGCTCTACTGATGTATCTCAATGAACACTGATACCGGATCTCACTT	1995
671	QY	671	PheAspThrAsnSerGlnGln-----LysSerLeuAlaSerGlnSerThrGlnLysGlu	688
1996	DB	1996	GAGGACAAACACATCAACACCAACCAACCAACCGGTGATAGCAACCAACGAGGAGGAGAA	2055
689	QY	689	LeuGlnGlyHisLeuAlaLeuThrThrGlnGluSerProHis-----	702
2056	DB	2056	-----CATTTTCTATGGGAGACGACGACGACTCTCATGACTTCTTCCCAATAAGT	2106
703	QY	703	-----ProGlnAsn-----PheGlnSerThrGlnGlnGlnThrHis	715
2107	DB	2107	CAGCCTTATGTGCTCTCGTTTGGGATCTTCTCTCTACCCCAAGAAAACCGA-----	2160
716	QY	716	LeuArgMetGluGluMetValThrIleAlaAlaSerSerProLeuPheSerHisHisAsp	735
2161	DB	2161	-----GCCAGCTCCATCCGGTTTTTCTGTGTCAACA	2190
736	QY	736	AspGlnTyrIleAlaGluAlaProThrGluHisThrGlyArgLysAspAlaLysLeu	755
2191	DB	2191	TGTCAGTGGCTTGGGAATTTGCCAATGTG-----GGTAATCAGAACCCCTTCCCATCC	2244
756	QY	756	ThrTyrGluGlnPheIysAlaThrThrArgAsnSerProAlaIaIaThrCysGlyAla---	774
2245	DB	2245	TCATTTTGGGTAATTACGTGCT-----TGTCATACTTCCCAAGTGT	2286
775	QY	775	-----GlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisValMetGly	792
2287	DB	2287	CCTAATCAATACAGA-----GAAGCTTCTCATCCAAATTGG	2322
793	QY	793	SerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAspArgTyrAlaGlu	812
2323	DB	2323	CCATCTTCC-----ATCATACCACACAGATCAATACAAGCCA	2361
813	QY	813	ArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGluAla	832
2362	DB	2362	GTTCCTTTAAATTAATCAAGTCAACAAAT-----CCGGTAGCTTCA-----	2412
833	QY	833	SerLysLeuCysArgArgAsnAlaGlyGlnValValLeuTyrProLysGlu-----	850
2413	DB	2413	TCTAACAAATGAAATACATGAACCTCAACTTTGTTGCTGCCAACGGAAGCAAAATGT	2472
850	QY	850	-----	850
2473	DB	2473	GGCCCTAATCCAGAAATTTCAATTTGGCTGCAACATGCTGCTGGGTTAGTAGTAGT	2532
851	QY	851	-----SerMetProAlaThrHisLeuLeuArg	859
2533	DB	2533	AGTAGGCCAATAGATAACTTTTCTAGTGAGAGCTCTATACCGCATTTGCATCTACTCAGC	2592
860	QY	860	MetMetAspProSerThrLeuAlaSerPhePro-----AsnTyrGlyThrSer-----	875
2593	DB	2593	CTTCTGGATCCTCGCTGAGGTCAACGACTCCCGCTGCAACACGGAACACATAAATTT	2652
876	QY	876	SerArgAsnGlnMetGluSerGlnLeuHisAsnSerGlnTyrAlaHisAsnGlnTyrLys	895
2653	DB	2653	ACTAAAGACATTTTCGCCAGCCCAACAGTCTCAAAAGAGTTTATAGAGCTTCAACAGGG	2712
896	QY	896	GlySerThrSerThrTyrGlySerAsnLeuAsnGlyLysIleProLeuThrPheGlu	915
2713	DB	2713	GACTCTAGTAAGTCAGCTACTCAACTAAG-----CAGATACCT-----TTTGA	2757
916	QY	916	AspLeuSerArgHisGlnLeuHisAspLeuHisArgProLeuArgPro---HisProArg	934
2758	DB	2758	CTATACAGCAAGAGATTCACACAGAGCCTTCCCGNAGAGTTTCCCATCATCTCCACCT	2817
935	QY	935	ValGlyValLeuGlySerLeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGly	954

Db	2818	ATTGGAGCG-----TCCTTCACTTTCAATTCCAAATGCTTCAATGGAGTCTCTCAT-----	2866
Qy	955	ThrGlnSerGlyTyrLysLeuGlyValSerThrGlyIleThrSerHisGlnMetAsnArg	974
Db	2866	-----CATCAAGAGAAAGAA-----	2880
Qy	975	LysGluHisPheGluAlaLeuAsnSerGlyMetPheSerAlaLysTrpAsnAlaLeuGln	994
Db	2881	ACCAAGAGAAAGACACC-----TTTGCTCCTCTTTTACATACTCATGAA	2925
Qy	995	LeuGlySerValSerSerSerAlaAsp-----PheLeuSerAlaAArgAsn	1009
Db	2926	ARGCCGGTGTTCAGCAAGCAGCAACCAAGCGAAGTTCAGCTGCTGGGAGCATCGAAT	2985
Qy	1010	SerIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHisProLeu-----	1026
Db	2986	TCC-----ATGATGCTTCTTGAAATTTTCAC	3012
Qy	1027	-----AspArgPheValArgGlnAsp-----	1033
Db	3013	ATGACGGATAAAGAAAGAAACAAAGAGAGAAAGACAGAGAGCTGCAATAACAAATGCTCT	3072
Qy	1034	-----IleCysIleThrAsnLysAsnProAla	1042
Db	3073	GCGGGACCTGTGAAGAACAGTTCTCGAACCATTTGTGTCAGCGTCAATGAACCCCTGCT	3132
Qy	1043	AspPheThrThrIleSerAsnAspAsnGluTyrMet	1054
Db	3133	GAITTCACCATCTCTGCACTCGGAATGTTTACATG	3168
RESULT 7			
LOCUS	AR368885	8648 bp	DNA
DEFINITION	Sequence 2 from patent US 6376751.		linear
ACCESSION	AR368885		
VERSION	AR368885.1	GI:34603269	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 8648)		
AUTHORS	Sung,Z.R., Aubert,D. and Chen,L.		
TITLE	Nucleic acids encoding EMF1 that control reproductive development in plants		
JOURNAL	Patent: US 6376751-A 23-APR-2002,		
FEATURES	Location/Qualifiers		
source	1..8648		
	/organism="unknown"		
	/mol_type="genomic DNA"		
ORIGIN			
Alignment Scores:			
Pred. No.:	4,07e-08	Length:	8648
Score:	298.00	Matches:	225
Percent Similarity:	35.53%	Conservative:	152
Best Local Similarity:	21.21%	Mismatches:	360
Query Match:	5.39%	Indels:	324
DB:	6	Gaps:	50
US-09-721-114-2 (1-1057) x AR368885 (1-8648)			
Qy	198	AlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAla	217
Db	5090	TCTAATATTGAGGACGACGACGAGAAAGCTTGTGAGTCCGAGCGAG-----GTAGGA	5140
Qy	218	AlaAsnValSerGluAsp-----AsnThrSerValAspValGlyAlaLeuProGlu	234
Db	5141	AACCAACAGATCTAAGAGAAAACCTAAACAAACCATCGATGATATTAGCAGCTGGAAGAG	5200
Qy	235	ValProGlnIleThrTrpHisIleGluValAsnGlyAla-----	247
Db	5201	AAACAAATGTGATCAGGTGCTTGACAACAGTTTCGGCTCATCTCTGAAATTCGTGTGGTGT	5260

QY	248	---AspGlnProProSerThrProLysLeuSerGluValValLeuLysArgAsnGluAsp	266
Db	5261	GAGGATACCACTTAAAGCAACCAAGCAATCAATAAGGCAATTCGGCGTCTGATGCAATGC	5320
QY	267	GluAsnGlyLysThrGluGluThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro	286
Db	5321	GATACCGGTTCATCAGAAAGTATAATCTTGCT	5353
QY	287	AsnProMetSerGlyLysGluArg	303
Db	5354	-----ATGAGTGGGTTCGAGCGTAGGAAATCTCGAAGGTTCTACTACTAGTGGTTG	5407
QY	304	ThrLysAspProLysProValSerGlyGlnLysCysGluGlnIleCysAsnGluProCys	323
Db	5408	CTTGGTAATACAAAACCAAGTGTGTAGTAACATCAGAAA	5449
QY	324	GluGluValValLeuLysArgSerLysSerLysArgLysThrAspLysLysLeuMet	343
Db	5450	GAAAGCTCTGCTTTGAAGAAGGAATCAGTTAGAGTTCGAAA	5503
QY	344	LysLysGlnGlnHis	360
Db	5504	CCTGAAAACAATATGTCAGCGGATATTGAGTACAAATGGTGGCAACCTCTGAAAATGCT	5563
QY	361	AlaLysLeuCys	364
Db	5564	TCCAAAAGTTGTGACTCTGATCAAGGTAATAGTGAATCAACTGATAGTGGTTTGACAGA	5623
QY	365	-----ArgArgLysProLysLysValArgLeuLeuSerGluLeuLeuLeu	379
Db	5624	ACTCCATTAAAGGTAAGCAGAGAAACGAAAGATTTCAGGTTTGTGACGAGTTGTACCA	5683
QY	380	AlaAsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspPro	399
Db	5684	TCACTTCTGTGTAACCTTCACAA	5737
QY	400	CysGluAspAspArgSerThr	408
Db	5738	AGTAAG-----AGATCAACTCTCGCGCACTCTTTATTCTCACTGGAACGATTTCTGTCT	5791
QY	409	ValProMetGluValSerMetAspIleProValSerAsnHisThrValGlyGluAspGly	428
Db	5792	TGTCTT-----CCGGTACTCAGAGAACA	5824
QY	429	LeuLysSerSerLysAsnLysThrLysArgLysTyrSerAspValValAspAspGlySer	448
Db	5825	CTCAGTTTACCAAGAAAGAGAACAAAGAG-----CCTGTAATCGAATAATGGGAAG	5875
QY	449	SerLeuMet-----AsnTrpLeuAsnGlyLysLys-----LysArgThrGly	462
Db	5876	AGCACTGATCATAGTTTATGTAACCGCAATTCGAGGAGTCAAGTTAACTCCGCATCTGCT	5935
QY	463	SerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysValThrPro	482
Db	5936	CCTTCCATGACACACAGTATCCCAACTCGAGACTTATTGAAATGGGAAAGGTTGGCGGT	5995
QY	483	ThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAsp	499
Db	5996	TTATTTCACAACCGTTTGGCTTCAGATGGATATTTCCAGAAAATATCTCTCAGGTTAAT	6055
QY	500	-----ThrAsnMetHisLysThrAspValCysGlnHisValSerGluLysSer	515
Db	6056	GATAAGCCGATACATCTTTGCAATTCGACAGCAATGATTATGTCAGGTCAAGAGACGG	6115
QY	516	ThrGlnArgCys-----SerSerLysGlyLysThrAlaGly-----Leu	528
Db	6116	GAACCAAACTGTCTTCGAGATTTTAGTCTCTCTTCTTAAATCCAGCTCAGGTGGTGGTTG	6175
QY	529	SerLysGly-----LysThr	533
Db	6176	AGAACTGGAGTAGATATTGTTGATCTTCAGAACCAACCAACCAATACAAACAGATCGTCT	6235
QY	534	HisSerAlaAlaSerThrLysTyrGlyGlyGluSerThrArgAsnGlyGlnAsnIleHis	553
Db	6236	TTCTCGAAGCTTGAAGCTAAGATACCCCTTCTTCTACTGAAGTTGCGGATTTATCTCGG	6295
QY	554	ValLeuSerAlaGluAspGln-CysGlnMetGluThrGluAsnSerValLeuSerHis	573
Db	6296	GTGCTGCAAAAGGTACAGCTTTGATCTATTTTCTTCAATCGCCATGCTTTCTCATTC	6355
QY	573	-----	573
Db	6356	TGTTTTCAACTGACACTATTTCTGTATATTTCCAAACTGGTGTCTTCACTGATGATGAT	6415
QY	574	-----Ala-LysValSerProAlaGluHisAspIleG	584
Db	6416	CAGGATGATAACGGGATGATTACTTGGCAGGATGCTTCTGTGTGCAGATGAAGGGGA	6475
QY	584	InIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysLysGlnLysLeuG	604
Db	6476	AGACTGTTATGCTTCCAAAGACATCATGAGCACCACCAAGAACCAAGTCACCATGAAGG	6535
QY	604	IuValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuL	624
Db	6536	AGACTACGACTGAGAGCAAAACAAAC--GATGATATTCCATGAGATAGTGGAGCTCA	6592
QY	624	euaLysAsnGlnHisGluArgGlnLeuMet-----ThrGluThrAspCysSerAspIleA	643
Db	6593	TGGCCAAAACCACTACGAGAGGTGCTTCCCGACAAAGAAAGAGATGTTAGCAACAAC	6652
QY	643	snArgIleGlnSerLysThrAlaAspAspCysValIleValAlaAlaLysAspG	663
Db	6653	AGCATCAACAGAAACAGCACACAAATCCAGATGCTCTACTGATGATGATCAATGAA	6712
QY	663	lySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGln-----LysSerLeuA	681
Db	6713	CCTAGATAACGGGATCTCACTTGAGGACCAACACACATCAAGACCAACCAACCGTGA	6772
QY	681	laSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGluSerP	701
Db	6773	GTAGCAACGCAAGGAGGAGNA-----CAITTTCTATGGGAGACAGCAGACT	6823
QY	701	roHis-----ProGlnAsn-----PheGlns	708
Db	6824	CTCATGACTTCTCCCAATAAGTCAGCCTTATGTGCTTCTCCGTTTGGGATCTTCTC	6883
QY	708	erThrGlnGluGlnThrHisLeuArgMetGluGluMetValThrIleAlaLysers	728
Db	6884	CTACCAAGAAACCGA-----GCCAGCTGCTG-----G	6907
QY	728	erProLeuPheSerHisHisAspGlnTyrIleAlaGluAlaProThrGluHisTrpG	748
Db	6908	CCATCCGTTTTCTGGTCACAACTGTGAGTGGCTGGGAATTTGCCAATGCTG-----	6961
QY	748	lyArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSerP	768
Db	6962	GTAATCAGAACCTTCTCCATCTCTTCCGGGTATTACGTGCT-----	7005
QY	768	roAlaAlaThrCysGlyAla-----GlnPheArgProGlyIleGlnAlaValAspL	785
Db	7006	--TGTGATCTTTCAGAGTGTCTCTAATCAATACAGA-----	7041
QY	785	euthrSerThrHisValMetGlySerSerSerAsnTyrAlaSerArgGlnProValIleA	805
Db	7042	--GAAGCTTCTCATCCCAATTTGGCCATCTCTCC-----ATGATAC	7078
QY	805	laProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProS	825
Db	7079	CACACAGAGTCAATACAGCCAGTCTTTTAAATATTAATCATCAGTCAACAAT--COGG	7135
QY	825	erThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValV	845
Db	7136	GTACGCTTTCA-----CAGGCACTTAACATGAATAATACATGGAACCTCACTTTGTG	7189
QY	845	allLeuTyrProLysGlu-----	850

Db 7190 CTGCCAACGGGAGCAAAATGTGGGCTTAATCCAGAAATTTTCATTTGGTGTGCAAAACATG 7249  
Qy 851 -----SerM 852  
Db 7250 CTGCTGGGGTTAGTAGTAGTAGTAGGCAATAGATACTTTCTAGTGAGAGCTCTA 7309  
Qy 852 etProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPhePro 870  
Db 7310 TACGGGATTCATCTACTCAGCTCTCGCTCTGAGTCAACAGCTCCCGCTG 7369  
Qy 971 --AsnTyrGlyThrSer-----SerArgAsnGlnMetGluSerGlnLeuHisAsnSerG 888  
Db 7370 ACCAACACGGGAACACTAAATTTACTAAAGACATTTTCCGCCAGCCCAACAGCTCCAAAG 7429  
Qy 888 lntYrAlaHisAsnGlnTyrLysGlySerThrSerThrSerThrSerThrSerThrSerThrSer 908  
Db 7430 AGTTTATAGAGCTTCAACACAGGGAGCTCTAGTAGTCAGCTACTCACTAAG----- 7482  
Qy 908 lylsilleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAsnLeuHisArgp 928  
Db 7483 --CAGATACCT-----TTTGATCTATACAGACAGATTCACACAGAGAGCTCCCGGA 7534  
Qy 928 roLeuArgPro--HisProArgValGlyValLeuGlySerLeuGlnLysGluLeuA 947  
Db 7535 AGATTTCCTCCCATCCTCCACCTATGAGAG-----TCTCATCTTCATTTCAAAATG 7588  
Qy 947 laAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSerThrGlyI 967  
Db 7589 CTTTCATGAGTCTCTCAT----- 7605  
Qy 967 leThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheS 987  
Db 7606 -----CATCAGAGAGAAACCAAGAGAGAAAGACAC-----TTTG 7642  
Qy 987 erAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAsp----- 1003  
Db 7643 CTCCTGTTTACATACTCATGAAAGCGGTGTTTGCACAGACGACGACCAAGCGAAGT 7702  
Qy 1004 -----PheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgLysGlyLysM 1022  
Db 7703 TCCAGCTGCTGGGAGCATCGAATTC-----A 7729  
Qy 1022 etValHisProLeu-----AspArgPheValArgGlnAsp----- 1033  
Db 7730 TGATGCTTCCTTTGAAATTCACATGACGATGAATAAGAAAGAAACAAAGAGAAAGACAG 7789  
Qy 1034 -----IleC 1035  
Db 7790 AGAGCTGCAATPACAATGCTCTCGGGACCTGTGAAGAACAGTTCGTGACCCCATTTGT 7849  
Qy 1035 ysilleThrAsnLysAsnProHisAspPheThrThrIleSerAsnAspAsnGluTyrMet 1054  
Db 7850 GCAGCGTCAATAGAAACCTGCTGATTCACCATTCCTGACCTGGAATGTTTACATG 7908  
RESULT 8  
AF319968 8648 bp DNA linear PLN 05-SEP-2001  
LOCUS Arabidopsis thaliana embryonic flower 1 (EMF1) gene, complete cds.  
DEFINITION Arabidopsis thaliana (thale cress)  
ACCESSION AF319968.1 GI:15430696  
VERSION AF319968.1  
KEYWORDS  
SOURCE  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsie.  
1 (bases 1 to 8648)  
REFERENCE  
AUTHORS Aubert,D., Chen,L., Moon,Y.H., Martin,D., Castle,L.A., Yang,C.H.  
and Sung,Z.R.  
TITLE EMF1, a novel protein involved in the control of shoot architecture  
and flowering in Arabidopsis  
JOURNAL Plant Cell 13 (8), 1865-1875 (2001)  
MEDLINE 21380417

11487698  
2 (bases 1 to 8648)  
REFERENCE  
AUTHORS Hubert,D., Chen,L. and Sung,Z.R.  
TITLE Direct Submission  
JOURNAL Submitted (07-NOV-2000) Plant and Biochemistry, Riscoe,  
Frederiksborgvej, Roskilde 4000, Denmark  
FEATURES  
Location/Qualifiers  
source  
1..8648  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3702"  
1..8648  
/chromosome="5"  
/gene="EMF1"  
1..3201  
/gene="EMF1"  
join(3202..3265,4160..4335,4448..4623,4704..4823,  
4903..4956,5046..6307,6448..8648)  
/gene="EMF1"  
/product="embryonic flower 1"  
join(4241..4335,4448..4623,4704..4823,4903..4956,  
5046..6307,6448..8031)  
/gene="EMF1"  
/codon\_start=1  
/product="embryonic flower 1"  
/protein\_id="AAK98528.1"  
/db\_xref="GI:15430697"  
/translation="MGSSIKINSISIDLAGAANEIDMVKDHFSMRGFAETREDLR  
KCPSEESVSVLQSYTLPTLSVFKRWKHCIMKIDIDAHGPPDCGLHNSKAIG  
NSVTSKSKFNSLTIDHEKKTIDADNAIEKVGVCENDDDQATVTFKCARGP  
MGASVRSKSKRLVSPQVGNRSKELANKPSMDISSWKEKONVQVDAVTFKSSBIAG  
VVEDTPPKATKHGIRGLMECDNGSSSINLMSGLORRKRKRVLLSGLGNTKTS  
GGSINRNSALKESVRGKRKLLLPENNYVSRILSTMGATSENASKSCDSQGNSES  
TSGEDRTFFKGRNRNFQVDFVPSLPCETISQEGIKHDADPSKSTPAHSLFTG  
VSDRLNGKRVGGLFNRLASDGYFRYLSQVNDKPTISLHLQNDYVRSADRP  
NDSVQPTQTERKLSLPKKTTPKVPIDNGKSTVISFNSGIDGQVNSHTGPSWNT  
CURPSSSSSSSGGWLRTGVIVDFRNHNHNRSSFSNKLRLYPPSSTEVADLRV  
LKDASGDRKGTVMQSHGAPRSQSHDRKETTTEBONDDIPMEIVELAKNOYE  
RGLPKEEVSNKQPSQTAHKSNAALIDINETYDNGISLENNTSRPPKPCSSNAR  
RHEHPMGQQSHDFPPIQSVVSPFGIFPTQENRASSIREFSGNCOMGLNLTVP  
GNQSPSPFRVLRACDTQCVQNFQREASHFIMFSSMLFPQSYKPSVLINQSTNP  
GTLQASNNNTNLFVAAKGKQGNPFESFGCKHAAGVSSSSSRSDIDNPFSSSS  
ITPALHLLSLDPLRSTTPADQKNTKTKRHPFANQKEFIELOTGSSKSAYSTK  
QIPDLYSKRFTQEPKSRKPPITPPTIGTSSLSFQNASWSPHHEKTKRDKTPAPVN  
THEKPVFASNDQAKFLIGASNNMLPLKFTMTDKKQKQKABSCNNNAGAPVN  
SSGPIVCSVNRNPADFTIPEPGNVYMLTGEHLKVRKRTTFKKKPAVCKQDAMKQTKPP  
VCPPTQNA"  
ORIGIN  
Alignment Scores:  
Pred. No.: 4,07e-08 Length: 8648  
Score: 298.00 Matches: 225  
Percent Similarity: 35.53% Conservative: 152  
Best Local Similarity: 21.21% Mismatches: 360  
Query Match: 5.39% Indels: 324  
DB: 8 Gaps: 50  
US-09-721-114-2 (1-1057) x AF319968 (1-8648)  
Qy 198 AlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAla 217  
Db 5090 TCTAATGTTAGGAGCAAGACAGAAAGCTGTGAGTCCGAGCAG-----GTAGGA 5140  
Qy 218 AlaAsnValSerGluAsp-----AsnThrSerValAspValGlyAlaLeuProGlu 234  
Db 5141 AACACAGATCTTAAGAAAACCTAAACAAACCTACGATGATATTAGCAGCTGGAAGAG 5200  
Qy 235 ValProGlnIleThrTrpHisIleGluValAsnGlyAla----- 247  
Db 5201 AAACAAAATGTCGATCAGCTGTCGACACAGCTTCGGCTCATCTCGAATTCGTGTGTTG 5260  
Qy 248 ---AspGlnProProSerThrProLysLeuSerGluValValLeuLysArgAsnGluAsp 266

5261 GAGGATACACCACTAAGGCAACCAAGATATCAAGGCAATTCGCGGTCTGATGAATGC 5320 Db  
267 GluAsnGlyThrGluThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro 286 Qy  
5321 GATACGGGTCTCAGAAAGATTAATCTTGCT----- 5353 Db  
287 AsnProMetSerGlyLysGluArg-----AspGlnValAlaGluGlnCysAsnLeu 303 Qy  
5354 -----ATCAGTGGGTTCGAGGTAGCAAAATCTCGCAAGTTCGCTACTCTAGTGAAGTTG 5407 Db  
304 ThrLysAspProLysProValSerGlyGlnLysCysGluGlnIleCysAsnGluProCys 323 Qy  
5408 CTTGGTAATACAAACCAAGTGTGTGTAGTAACATCAGAAAA----- 5449 Db  
324 GluGluValValLeuLysArgSerLysSerLysArgLysThrAspLysLysLeuMet 343 Qy  
5450 GAAGAGTCTGCTTTGAAGAGGAATCAGTTAGAGTTCGAAAA-----AGAAAGTTGTTA 5503 Db  
344 LysLysGlnGlnHis---SerLysLysArgThrAlaGlnAlaAspValSerAsp----- 360 Qy  
5504 CTTGAAAAACAATTATGACGCCGATATTGAGTACAAATGGTGGCAACCTCTGAAAAATGCT 5563 Db  
361 AlaLysLeuCys----- 364 Qy  
5564 TCCAAAAGTTGTGACTCTGATCAAGGTAATAGTGAATCAACTGATAGTGGGTTTGACAGA 5623 Db  
365 -----ArgArgLysProLysLysValArgLeuLeuSerGluIleIleAsn 379 Qy  
5624 ACTCCATTTAAGGTAAGCAGAGAAACAGAGATTTTCAGGTGTGACAGTTGTGACCA 5683 Db  
380 AlaAsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspPro 399 Qy  
5684 TCACCTCCCTGTGAACTTCACAA-----GAAGGTATCAAGGATGATGATGATGATGAT 5737 Db  
400 CysGluAspAspArgSerThr-----IlePro 408 Qy  
5738 AGTAAG-----AGATCAACTCTCGCACTCTTTTATCTACTGGAACAGATCTGTCTCT 5791 Db  
409 ValProMetGluValSerMetAspIleProValSerAsnHisThrValGlyGluAspGly 428 Qy  
5792 TGTCTCT-----CCGGTACTCAGAGACA-----GAGAGGAAG 5824 Db  
429 LeuLysSerSerLysAsnLysThrLysArgLysTySerAspValValAspAspGlySer 448 Qy  
5825 CTCAGTTTACCCAGAGAGAGACAAAGAA-----CCTGTATTCGATATATGGGAAG 5875 Db  
449 SerLeuMet-----AsnTrpLeuAsnGlyLysLys-----LysArgThrGly 462 Qy  
5876 AGCACTGTATCAGTTTGTAGTACGGCATTTGATGGAAGTCAAGTCAATCTGCACTACTGTT 5935 Db  
463 SerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysValThrPro 482 Qy  
5936 CTTCCATGAACACAGATATCCAAACTCGAGACTTATTGAATGGGAAAGGGTGGCGGT 5995 Db  
483 ThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAsp----- 499 Qy  
5996 TTATTTCACCAACCGTTTGGCTTCAGATGGATATTTTCAGAAAAATATCTCTCTCAGTTAAT 6055 Db  
500 -----ThrAsnMetHisLysThrAspValCysGlnHisValSerGluIleSer 515 Qy  
6056 GATAAGCCGATAACATCTTTCGATTTCCAGACAAATGATTATGTGAGGTCAAGAGACGCG 6115 Db  
516 ThrGlnArgCys-----SerSerLysGlyLysThrAlaGly---Leu 528 Qy  
6116 GAACCAAACTGTCTTCGAGATTTTGTGTCCTCTCTTAATCCAGCTCAGGTGGATGTTG 6175 Db  
529 SerLysGly-----LysThr 533 Qy  
6176 AGAACTGGAGTAGATATTGTGACTTCAGAAACAACAACCAACACATACAAACAGATCGTCT 6235 Db  
534 HisSerAlaAlaSerThrLysTyGlyGluSerThrArgAsnGlyGlnAsnIleHis 553 Qy  
6236 TTCTCGAACTTGAAGTAAAGATACCCCTCTCTTCTACTGAAAGTGGCGATTTATCTCGG 6295 Db

554 ValLeuSerAlaGluAspGln-CysGlnMetGluThrGluAsnSerValLeuSerHisSe 573 Qy  
6296 GTGCTGCAAAAGGTACAAGCTTTGCTATCTTTTCTCAATCGCCATGCTTCTTCATTC 6355 Db  
573 K----- 573 Qy  
6356 TGTTTTCCAACTTGCACTATCTTGTATATTCCAAACCTGTGTGCTTCAGTGATGGAT 6415 Db  
574 -----Ala-LysValSerProAlaGluHisAspIleG 584 Qy  
6416 CAGCATGATACCGGATTTGATTTACTTTGGCAGGATGCTTCTGTGTCAGATAGAAGGGA 6475 Db  
584 InIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysLysLysLeuG 604 Qy  
6476 AGACTGTATGTCCTCAAGACATCATGAGCAGCACCAAGAGCCAAAGTCACCATAGAAAGG 6535 Db  
604 LuValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeu 624 Qy  
6536 AGACTACGACTCAAGAGCAAAACAAC---GATGATATTCCAAATGGAGATAGTGGAGCTCA 6592 Db  
624 euAlaLysAsnGlnHisGluArgGlnLeuMet---ThrGluThrAspCysSerAspIleA 643 Qy  
6593 TGGCCAAAACACAGTACGAGAGGTGCTTCCCGCAAGAGAGAGATGTTAGCAACAAC 6652 Db  
643 snArgIleGlnSerLysThrAlaAspAspAspCysValIleValAlaAlaLysAspG 663 Qy  
6653 AGCCATCAACAAGAAACAGACACAAATCCAAAGATGCTCTACTGATTCATCAATGAA 6712 Db  
663 LysSerAspTyAlaSerSerValPheAspThrAsnSerGlnGln-----LysSerLeuA 681 Qy  
6713 CCTAGATTAACGGGATCTCATCTTGGAGACAACAACATCAAGACCACCAAAACCGTGT 6772 Db  
681 laSerGlnSerThrGlnLysGluGlnGlyHisLeuAlaLeuThrThrGlnGluSerP 701 Qy  
6773 GTAGCAACGAGGAGGAGAA-----CATTTCTTATGGGAAGACAGCAGAACT 6823 Db  
701 roHis-----ProGlnAsn-----PheGlns 708 Qy  
6824 CTCATGACTTCTCCCAATAAGTCAGCCTTATGTCCTCTCCGTTTGGGATCTTCTC 6883 Db  
708 erThrGlnGluGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaLysS 728 Qy  
6884 CTACCCAGAAACCGA-----GCCAGCT 6907 Db  
728 erProLeuPheSerHisAspAspGlnTyIleAlaGluAlaProThrGluHisTrp 748 Qy  
6908 CCATCCGGTTTTCTGCTCACAACCTGTCAGTGGCTTGGGAATTTGCCAACTGTG-----G 6961 Db  
748 lYArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSerP 768 Qy  
6962 GTAATCAGAACCTTCTCCATCTCTCATTTCCGGTATTACGTGCT----- 7005 Db  
768 roAlaAlaThrCysGlyAla-----GlnPheArgProGlyIleGlnAlaValAspL 785 Qy  
7006 --TGTGATACTTGCAGAGGTTCCTTAATCAATACAGA----- 7041 Db  
785 euThrSerThrHisValMetGlySerSerSerSerSerSerSerSerSerSerSerSer 805 Qy  
7042 --GAGCTTCTCATCCAAATTTGCCATCTTCC-----ATGATAC 7078 Db  
805 laProLeuAspArgTyAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProS 825 Qy  
7079 CACCACAGAGTCAATACAGCCAGCTTCTTTAAATTAATTAATCAGTCAACAAAT---CCGG 7135 Db  
825 erThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValV 845 Qy  
7136 GTAAGCTTTTCA-----CAGGCATCTAAATGAATATACATGGAACCTCAACTTTGTTG 7189 Db  
845 allLeuTyProLysGlu----- 850 Qy  
7190 CTGCCAACGGAGCAAAATATGGGCTAATCCAGAAATTTTCAATTTGCTGCAACATG 7249 Db

```

QY 851 -----SerM 852
Db 7250 CTCGCGGGTTAGTAGTAGTAGAGCCAAATAGTAGTAATCTTCTAGTAGGAGCTCTA 7309
QY 852 etProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPhePro----- 870
Db 7310 TACCGGCATTGCATCTACTACGCGCTCTGATCTCTGATCTCTGATCTCAACGCTCCCGCTG 7369
QY 871 --AsnTyrGlyThrSer-----SerArgAsnGlnMetGluSerGlnLeuHisAsnSerG 888
Db 7370 ACCAAGCAGGAACACTAATTTACTTAAAGACATTTCCGCCAGCCCAACCACTCCANAG 7429
QY 888 lnTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerThrSerThrSerThrSerThr 908
Db 7430 AGTTTATAGAGCTTCAACACGGGAGCTCTAGTAGTCAAGTCAAGCTCAACTAAG----- 7482
QY 908 lylsileProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgP 928
Db 7483 --CAGATACCT-----TTTGATCTATACAGCAGAGAGATTACACAGAGCCCTCCCGGA 7534
QY 928 roLeuArgPro---HisProArgValGlyValLeuGlySerLeuLeuGlnLysGluLeuA 947
Db 7535 AGAGTTTCCCATCCTCCATCCACTATTGGGACG-----TCTTCACTTTCATTTCAAAATG 7588
QY 947 laAsnTyrSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSerThrGlyI 967
Db 7589 CTTCTGAGGAGCTCTCAT----- 7605
QY 967 leThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheS 987
Db 7606 -----CATCAGAGAGAAACCAAGAGAAAGACACC-----TTTG 7642
QY 987 erAlaIysTyrAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAsp----- 1003
Db 7643 CTCCTGTTTCAATACTCATGAAAGCCGGTGTTCGACGACGACCAACGACGCGAAGT 7702
QY 1004 -----PheLeuSerAlaArgAsnSerIleAlaGlnSerTyrThrArgGlyLysGlyLysM 1022
Db 7703 TCACGCTGCTGGGAGCTCGAATCC-----A 7729
QY 1022 etValHisProLeu-----AspArgPheValArgGlnAsp----- 1033
Db 7730 TGATGCTCTCTTTCGAAATTTTCATGATGACGATGATGATGATGATGATGATGATGATG 7789
QY 1034 ----- 1035
Db 7790 AGAGCTGCAATACATGCTCTGCGGACCTGTGAGAACACATCTCTGACCCATTTGTGT 7849
QY 1035 ysileThrAsnLysAsnProAlaAspPheThrThrIleSerAsnAspAsnGluTyrMet 1054
Db 7850 GCAGCGTCAATAGAACCTGCTGATTTACCATTTCCATTTCCATTTCCATTTCCATTT 7908
RESULT 9
AR368884
LOCUS
DEFINITION Sequence 1 from patent US 6376751.
AUTHORS AR368884
ACCESSION AR368884
VERSION AR368884.1 GI:34603268
KEYWORDS
SOURCE
ORGANISM Unknown.
Unclassified.
REFERENCE
1 (bases 1 to 17341)
AUTHORS Sung,Z.R., Aubert,D. and Chen,L.
TITLE Nucleic acids encoding EMF1 that control reproductive development
in plants
JOURNAL Patent: US 6376751-A 1 23-APR-2002;
FEATURES Location/Qualifiers
1..17341
/mol_type="genomic DNA"
ORIGIN

```

```

Alignment Scores:
Pred. No.: 9,7e-08 Length: 17341
Score: 298.00 Matches: 225
Percent Similarity: 35.53% Conservative: 152
Best Local Similarity: 21.21% Mismatches: 360
Query Match: 324
DB: 50
US-09-721-114-2 (1-1057) x AR368884 (1-17341)
QY 198 AlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAla 217
Db 5090 TCTAATGTTAGGAGCAGAGCAGAAAGCTTGTAGTCCGGACAG-----GTAGGA 5140
QY 218 AlaAsnValSerGluAsp-----AenThrSerValAspValGlyAlaLeuProGlu 234
Db 5141 AACCAACAGATCTTAAAGAAAACTAAACCAAAACCATCGATGATATTAGCAGCTCGAAGAG 5200
QY 235 ValProGlnIleThrTrpHisIleGluValAsnGlyAla----- 247
Db 5201 AAACAAAATGTGGATCAGGCTGTGACACGTTCCGGCTCATCTGAAATCTCTGGTGGTGT 5260
QY 248 ---AspGlnProProSerThrProLysLeuSerGluValValLeuLysArgAsnGluAsp 266
Db 5261 GAGGATACACCACTTAAGCAACCAAGCAATCATAAAGCATTCGGGCTCTGATGGAATGC 5320
QY 267 GluAsnGlyLysThrGluGluThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro 286
Db 5321 GATAACGGGCTCATCAGAAAGTATAAATCTTGTCT----- 5353
QY 287 AsnProMetSerGlyLysGluArg-----AspGlnValAlaGluGlnCysAsnLeu 303
Db 5354 -----ATGAGTGGTTCGACGGTAGGAATCTCGCAAGTTTCGTCTACTCAGTGAGTTG 5407
QY 304 ThrLysAspProLysProValSerGlyGlnLysCysGluGlnIleCysAsnGluProCys 323
Db 5408 CTTGTAATACAAAAACCAACGAGTGTGTGATTAACATACAGAAAA----- 5449
QY 324 GluGluValValLeuLysArgSerSerLysSerLysArgLysThrAspLysLysLeuMet 343
Db 5450 GAAGAGTCTGCTTTGAAGAGGAATCAGTTAGAGTTCGAAAA-----AGAAGTTGTTA 5503
QY 344 LysLysGlnGlnHis---SerLysLysArgThrAlaGlnAlaAspValSerAsp 360
Db 5504 CCTGAAACAAATATTGTGACGGGATATTGATACATATGGGTGCAACCTCTGAAATGCT 5563
QY 361 AlaLysLeuCys----- 364
Db 5564 TCCAAAAGTTGTGACTCTGATCAAGGTATAGTGAATCAACTGATAGTGGGTTTGACAGA 5623
QY 365 -----ArgArgLysProLysLysValArgLeuLeuSerGluIleLeuAsn 379
Db 5624 ACTCCATTTAAGGTAAGCAGAGAACAGAGATTTCAGTTGTTGACGAGTTTGATACCA 5683
QY 380 AlaAsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspPro 399
Db 5684 TCACCTTCTCTGTGAAACCTTCACA-----GAAGGTATCAAGGAGCATGATCAGATCCT 5737
QY 400 CysGluAspAspArgSerThr-----llePro 408
Db 5738 AGTAG-----AGATCAACTCTCGGCACTCTTTATTCACCTGGAACGATTTCTGCTCT 5791
QY 409 ValProMetGluValSerMetAspIleProValSerAsnHisThrValGlyGluAspGly 428
Db 5792 TGTCTCT-----COGGGTACTCAGAGAACA-----GAGAGGAAG 5824
QY 429 LeuLysSerSerLysAsnLysThrLysArgLysTyrSerAspValValAspAspGlySer 448
Db 5825 CTCAGTTTACCCAGAGAAGAAAGCAAGAG-----CCTGTATCATTAATGGGAAG 5875
QY 449 SerLeuMet-----AsnTrpLeuAsnGlyLysLys-----LysArgThrGly 462
Db 5876 AGCAGCTGTGATCAGTTTATTAGTAACGGCATTTGATGCAAGTCAAGTTAACTCGCATACTGCT 5935

```

Qy	463	SerValHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysValThrPro	482
Db	5936	CCTTCATGAACACAGTATCCCAAACTCGAGACTTATTGAATGGGAAAGGGTGGCGGT	5995
Qy	483	ThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAsp	499
Db	5996	TTATTTGCACACCGTTGGCTTCAGATGGGATATTCGAAAAATATCTCTCAGGTTAAT	6055
Qy	500	-----ThrAsnMetHisLysThrAspValCysGlnHisValSerGluIleSer	515
Db	6056	GATAAGCCGATAACATCTTTTGCATTTGCAAGACAATGATTATGTGAGGTCAAGAGACGG	6115
Qy	516	ThrGluArgCys-----SerSerLysGlyLysThrAlaGly---Leu	528
Db	6116	GAACCAACTGTCTTCGAGATTTAGTTCCTCTCTAAATCCAGCTCAGGTGGATGGTTG	6175
Qy	529	SerLysGly-----LysThr	533
Db	6176	AGAACTGGAGTAGATATTGTTGACTTCAGAAACAAACACCAATACAAACAGATCGTCT	6235
Qy	534	HisSerAlaAlaSerThrLysThrGlyGlyLysSerThrArgAsnGlyGlnAsnIleHis	553
Db	6236	TTCTCGAAGCTGAAGCTAAGTACATCCGCCCTCTCTACTGAAGTTGGCGATTTATCTCGG	6295
Qy	554	ValLeuSerAlaGluAspGln-CysGlnMetGluThrGluAsnSerValLeuSerHisSe	573
Db	6296	GTGCTGCAAAAGGTACAAGCTTGTGATCTATTTTCTCAATGCGCATGCTTCTCATTC	6355
Qy	573	r-----	573
Db	6356	TGTTTTCCAACTGCACTATCTGTATATTCGAACTGGTGTCTCAGTGGATGGAT	6415
Qy	574	-----Ala-LysValSerProAlaGluHisAspIleG	584
Db	6416	CAGGATGATAACGGGATTGATTACTTGGSCAGGATGCTTCTGGTGCAGATAGAAAGGGGA	6475
Qy	584	InIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysGlnLysLeuG	604
Db	6476	AGACTGTTATGTGTCCAAGAACATCATGAGGACACACAGAACCAAGTCAGATGAAAGG	6535
Qy	604	IuValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuL	624
Db	6536	AGACTACGCTGAAGACCAAAACAAC--CATGATATTCCAATGGAGATAGTGGAGTCA	6592
Qy	624	eulAlaLysAsnGlnHisGluArgGlnLeuMet---ThrGluThrAspCysSerAspIleA	643
Db	6593	TGGCCAAAAACCAATACGAGAGGTGCTCTCCGACAAAGAAAGAGATGTTAGCAACAAC	6652
Qy	643	snArgIleGlnSerLysThr-ThrAlaAspAspCysValIleValAlaAlaLysAspG	663
Db	6653	AGCATCACAAAGAACACACAAATCCAGATGCTCTACTGATGATCTCAATGANA	6712
Qy	663	lySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGln-----LysSerLeuA	681
Db	6713	CCTACGATAACGGGATCTCACTGAGGACACACACATCAAGACCAACCAACCGGTGA	6772
Qy	681	laserGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThr-ThrGlnGluSerP	701
Db	6773	GTACACGCAAGGAGGGAAGAA-----CATTTTCTATGGGAAGACACGAGACT	6823
Qy	701	roHis-----ProGlnAsn-----PheGln	708
Db	6824	CTCATGACTTCTTCCCAATAAGTCAGCGCTTATGTGCGCTTCTCCGTTTGGGATCTTTCCTC	6883
Qy	708	erThrGlnGluGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerS	728
Db	6884	CTACCCCAAGAAACCGA-----	6907
Qy	728	erProLeuPheSerHisHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTrpG	748
Db	6908	CCATCCGGTTTCTGGTCAACATGTGATGGCTTGGCAATTTGCCAACTGTG-----G	6961

Qy	748	lyArgIysAspAlaLysLysLeuThrTrpGlnPheLysAlaThrThrArgAsnSerP	768
Db	6962	GTAAATCAGAACCCCTCTCCATCCTCATCTTTGGGTATACGTGCT	7005
Qy	768	roAlaAlaThrCysGlyAla-----GlnPheArgProGlyIleGlnAlaValAspL	785
Db	7006	--TGTGATACTTCCAGAGTGTCTCTTAATCAATACAGA	7041
Qy	785	euThrSerThrHisValMetGlySerSerSerAsnTyrTyrAlaSerArgGlnProValIleA	805
Db	7042	--GAAGCTTCTCATCAAAATTTGGCCATCTTC	7078
Qy	805	laProLeuAspArgTyrTyrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProS	825
Db	7079	CACCACAGACTCAATACAGCCAGTGTCTTTAAATATTAATCAGTCACCAAT	7135
Qy	825	erThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValV	845
Db	7136	GTACGCTTTCA-----CAGGCATCTAACAATGAAATAATACATGGAACCTCACTTTGTTG	7189
Qy	845	alleuTyrProLysGlu-----	850
Db	7190	CTGCCAACGGGAGACAAANAATGTGGCCTTAATCCAGAAATTTTCATTTGGCTGCAACATG	7249
Qy	851	-----SerM	852
Db	7250	CTGCTGGGTAGTAGTAGTAGTAGGCAATAGATAACTTTCTAGTAGCAGAGCTA	7309
Qy	852	etProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPhePro	870
Db	7310	TACGGGATTTGCATCTACTCAGCGTCTTGATCTCTCGCTCAGGTCAACGACTCCCGCTG	7369
Qy	871	--AsnTyrGlyThrSer-----SerArgAsnGlnMetGluSerGlnLeuHisAsnSerG	888
Db	7370	ACCAACACGGAGACACTAAATTTACTTAAAGACATTTCCGCCAGCCACCAGTCCCAAG	7429
Qy	888	lnTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsnG	908
Db	7430	AGTTATAGAGCTTCAACACGGGAGCTCTAGTAAGTCAGCCTACTCAACTAAG	7482
Qy	908	lyLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgP	928
Db	7483	--CAGATACCT-----TTTGATCTATACAGCAGCAGATTCAACAGAGCCTTCCCGA	7534
Qy	928	roLeuArgPro---HisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIleA	947
Db	7535	AGAGTTCCTCCCATCCTCCACTATTGGGAGC-----TCTTCACCTTCATTTCAAAATG	7588
Qy	947	laAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSerThrClyI	967
Db	7589	CTTCATGGAGTCTCAT	7605
Qy	967	leThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheS	987
Db	7606	-----CATCAGACAGAAACCAACAGAGAAAGACACC-----TTTG	7642
Qy	987	erAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAsp-----	1003
Db	7643	CTCCTGTTTCAATATCTCATGAAAACCGGTGTTTGCAAGCAGCAACGACCAAGCGAAGT	7702
Qy	1004	-----PheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysM	1022
Db	7703	TCCAGCTGCTGGGAGCATCGAATTC	7729
Qy	1022	etValHisProLeu-----AspArgPheValArgGlnAsp-----	1033
Db	7730	TGATGCTCTCTTGAATTTTCATCAGCGATATAAGAAAAAGAAACAAAGAGAAAAGCAG	7789
Qy	1034	-----IleC	1035
Db	7790	AGAGCTGCAATACAAATGCTCTGCGGGACCTGTGAGAACACAGTCTGAGCCCATGTGTG	7849
Qy	1035	ysIleThrAsnLysAsnProAlaAspPheThrThrIleSerAsnAspAsnGluTyrMet	1054

Db 7850 GCACGCTCATAGAAACCCCTGCTGATTCACCAATCTCTGAACTGGGAATGTTTACATG 7908

RESULT 10  
 ATP15N18 101665 bp DNA linear PLN 13-APR-2000  
 LOCUS Arabidopsis thaliana DNA chromosome 5, BAC clone F15N18 (BSSA project).

DEFINITION  
 Arabidopsis thaliana (thale cress)

ACCESSION  
 AL163815.1 GI:75733398

KEYWORDS  
 Arabidopsis thaliana

SOURCE  
 Arabidopsis thaliana

ORGANISM  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 101665)

REFERENCE  
 Bevan, M., Hilbert, H., Braun, M., Hweizer, E., Brandt, A., Duesterhoeft, A., Bancroft, I., Mewes, H. W., Rudd, S., Lemcke, K. and Mayer, K. F. X.  
 Unpublished

JOURNAL  
 2 (bases 1 to 101665)

REFERENCE  
 EU Arabidopsis sequencing project.

AUTHORS  
 Direct Submission

TITLE  
 Submitted (13-APR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de

JOURNAL  
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UU Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk

COMMENT  
 Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES  
 Location/Qualifiers

source  
 1..101665  
 /organism="Arabidopsis thaliana"  
 /mol\_type="Genomic DNA"  
 /variety="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="5"  
 6461..8797  
 /gene="F15N18\_10"  
 join(6461..6536,7295..7792,8271..8797)  
 /gene="F15N18\_10"  
 /note="similarity to predicted proteins in castor bean, Arabidopsis thaliana and alfalfa."  
 /codon\_start=1  
 /product="putative protein"  
 /protein\_id="CA887702.1"  
 /db\_xref="GI:7573399"  
 /db\_xref="SPTREMBL:O9LYE7"  
 /translation="MKGSGLSFLFVLLIATITVSCDGLPNEDFELGPKPDMKGQMYVINKAIPSWELSGPVEYIKSGQGDMLLVVPACKFAIRLGNBASIKQRLNVTKGTVISLTPSAARTCAQDERLINISVAPDGVPIQTVYSSGMDLYANAFQASNVARIVIHNGREEDPACGLIDGVAIKALYPRPTNKLKNGGPEGPVYLPNATTCGLVPPFIEDHSSPLANNVESLKAIKYVDVEHFSVQQRRAVELVAGRESALQAVARTVVGKTYLSFVAGNDANACGSGSVVEAFAGKDTLKVYPYESKGGKGFKKASLRFVAVSTRTRTMFYSTFTSMRSDFDSSLCGPPVDDVKLLSARKP"  
 6461..6536  
 /gene="F15N18\_10"  
 /number=1  
 6537..7294  
 /gene="F15N18\_10"  
 /number=1  
 7295..7792  
 /gene="F15N18\_10"  
 /number=2  
 7793..8270  
 /gene="F15N18\_10"  
 /number=2  
 8271..8797  
 /gene="F15N18\_10"  
 /number=3

gene  
 cds  
 exon  
 intron  
 exon  
 intron  
 exon





QY 516 ThrGlnArgCys-----SerSerLysGlyLysThrAlaGly---Leu 528  
DB 60821 GAACCAACGTGCTTCGAGATTTTTCAGTCTCTTCTAAATCCAGCTCAGGTGGATGGTGG 60880  
QY 529 SerLysGly-----LysThr 533  
DB 60881 AGAAGCTGGAGTAGATATTGTTGACTTCAGAAACAAACACCAATACAAACAGATCGTCT 60940  
QY 534 HisSerAlaAlaSerThrLysThrGlyGlyLysThrArgAsnGlyGlnAsnIleHis 553  
DB 60941 TTCTCGAAGCTGAAGCTAAGATACCCCTTCTTCTACTAGTTGCGGATTTATCTCGG 61000  
QY 554 ValLeuSerAlaGluAspGln-CysGlnMetGluThrGluAsnSerValLeuSerHis 573  
DB 61001 GTGCTGAAAAGGTACAAAGCTTTGCACTATTCTTCTCAATGCGCATCTTCTCATTC 61060  
QY 573 r----- 573  
DB 61061 TGTTTTCCAACTTGCACCTATTCTTGATATTCAAAAGTGGTGTCTTCAGTGGATGGAT 61120  
QY 574 -----Ala-LysValSerProAlaGluHisAspIleG 584  
DB 61121 CAGGATGATAACGGGATTTGATTTACTTTGGCAGGATGCTTCTGTCGATAGAAAGGGGA 61180  
QY 584 InIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysLysGlnLysLeuG 604  
DB 61181 AGACTGTTATGGTCCAAAGACATCATGAGCAGCACCAGAACCAAGTCACGATAGAAAGG 61240  
QY 604 LuValThrArgGluLysGlnThrMetIleAspIleProMetAspIleValGluLeuL 624  
DB 61241 AGACTAGGACTGAAGAGCAAAACAAAC---GATGATATTCCAAATGGAGATAGTGAGCTCA 61297  
QY 624 euAlaLysAsnGlnHisGluArgGlnLeuMet---ThrGluThrAspCysSerAspIleA 643  
DB 61298 TGGCCAAAACACAGTAGCAGAGGTGCTTCCCGCAAGAAAGAGAGATGTTAGCAACAAAC 61357  
QY 643 snArgIleGlnSerLysThrAlaAlaAspAspCysValIleValAlaAlaLysAspG 663  
DB 61358 AGCCATCACAAGAACACACACAAATCCAGAAATGCTCTACTGATGATCTCAATGAAA 61417  
QY 663 LysSerAspTyrAlaSerSerValPheAspThrAsnSerGlnGln-----LysSerLeuA 681  
DB 61418 CCTACGATAACGGGATCTCACTTGGAGACAAACACATCAAGACCAACCAACCGGTGA 61477  
QY 681 LaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrGlnGluSerP 701  
DB 61478 GTACCAAGCAGGAGGGAAGAA-----CATTTTCTATGGGAGACAGCAGAACT 61528  
QY 701 roHis-----ProGlnAsn-----PheGlns 708  
DB 61529 CTCATGACTTCTCCCAATAAGTCAGCCTTATGTCCTTCTCCGTTTGGGATCTTCTC 61588  
QY 708 erThrGlnGluGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerS 728  
DB 61589 CTACCCCAAGAAACCGA-----GCCAGCT 61612  
QY 728 erProLeuPheSerHisHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTrpG 748  
DB 61613 CCATCCGGTTTCTGGTCACAACTGTCTGAGTGGTGGGAATTTGCCAACTGTG-----G 61666  
QY 748 LysArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrArgAsnSerP 768  
DB 61667 GTAATCAGAACCTTCTCCATCTCTCATTTCCGGTATTACGTGCT----- 61710  
QY 768 roAlaAlaThrCysGlyAla-----GlnPheArgProGlyIleGlnAlaValAspL 785  
DB 61711 --TGATGACTTGCAGAGTGTCTTAATCAATACAG----- 61746  
QY 785 euThrSerThrHisValMetGlySerSerSerAsnTyrAlaSerArgGlnProValIleA 805  
DB 61747 --GAAGCTTCTCATCCAATTTGGCCATCTTCC-----ATGATAC 61783  
QY 805 laProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProS 825

DB 61784 CACCACAGAGTCAATACAGCCAGCTTTCTTTAAATATTAAATCAGTCAACAAAT---CCGG 61840  
QY 825 erThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValV 845  
DB 61841 GTACGCTTTCA-----CAGGCATCTAACAAATGAAATACATGGAACCTCAACTTTGTTG 61894  
QY 845 allLeuTyrProLysGlu----- 850  
DB 61895 GTGCCAACCGGGAAGCAAAATGTGGCCCTAATCCAGAAATTTTCATTGGCTGCAAAACATG 61954  
QY 851 -----SerM 852  
DB 61955 CTGCTGGGTTAGTAGTAGTAGTAGTAGGCAATAGATAACTTTTCTAGTGAGAGCTCTA 62014  
QY 852 etProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPhePro----- 870  
DB 62015 TACCGGCATTTGATCTACTCTCAGCCTTCTGGATCTCTGCTCAGAGTCAACGACTCCCGCTG 62074  
QY 871 --AsnTyrGlyThrSer-----SerArgAsnGlnMetGluSerGlnLeuHisAsnSerG 888  
DB 62075 ACCAACCGGGAACACATAAATTTACTTAAAGACATTTTCGGCCAGCCACACGATCCAAAG 62134  
QY 888 InTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsnG 908  
DB 62135 AGTTTATAGAGCTTCAACACAGGGGACTCTAGTAAGTCAAGCTACTCAACTAAG----- 62187  
QY 908 LysLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgP 928  
DB 62188 --CAGATACCT-----TTGATCTTATACAGCAAGAGATTCACAAAGACCTTCCCGGA 62239  
QY 928 roLeuArgPro---HisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIleA 947  
DB 62240 AGAGTTTTCCTCCACTCCTCCACTATTGGGACG-----TCTTCACTTCTATTCAAAATG 62293  
QY 947 laAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSerThrGlyI 967  
DB 62294 CTTCATGGAGTCTCTCAT----- 62310  
QY 967 leThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheS 987  
DB 62311 -----CATCAGAGAGAAACCAACAGAAAGACACC-----TTTG 62347  
QY 987 erAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAsp----- 1003  
DB 62348 CTCTCTGTTTACATACTCATGAAAGCCGTTGTTTGCAGACGACCAACGACGACGAAAGT 62407  
QY 1004 -----PheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysM 1022  
DB 62408 TCCAGCTGCTGGAGCATCGAATTC-----A 62434  
QY 1022 etValHisProLeu-----AspArgPheValArgGlnAsp----- 1033  
DB 62435 TGATGCTTCTTTGAAATTTTCATGACGATAAAGAAAGAAACAAAGAGAGAAAGCAG 62494  
QY 1034 -----IleC 1035  
DB 62495 AGAGCTGCAATACATGCTCTGGGGACCTGTGAAGAACAGTCTTGGAACCAITGTGT 62554  
QY 1035 ysIleThrAsnLysAsnProAlaAspPheThrThrIleSerAsnAspAsnGluTyrMet 1054  
DB 62555 GCAGCGCTCATAGAAACCTGCTGATTTCCATTCCTCACTGACCTGGGATGTTTACATG 62613  
RESULT 11  
MUSCVCOLORA  
LOCUS 7036 bp mRNA linear ROD 20-OCT-1999  
DEFINITION Mus musculus cyclophilin-related protein mRNA, complete cds.  
ACCESSION L04289  
VERSION L04289.2 GI:6073857  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae, Murinae; Mus. 1 (bases 1 to 7036) Anderson,S.K., Gallinger,S., Roder,J., Frey,J., Young,H.A. and Ortaldo,J.R.
TITLE	A cyclophilin-related protein involved in the function of natural killer cells
JOURNAL MEDLINE PIUMED	Proc. Natl. Acad. Sci. U.S.A. 90 (2), 542-546 (1993) 93133924 8421688
REFERENCE AUTHORS	Anderson,S. Direct Submission
TITLE	Direct Submission
JOURNAL	Submitted (27-Apr-1993) Lab of Experimental Immunology, NCI-FCRDC, Building 560, Room 31-93, Frederick, MD 21702-1201, USA
REFERENCE AUTHORS	Anderson,S. 3 (bases 1 to 7036)
TITLE	Direct Submission
JOURNAL	Submitted (20-Oct-1999) Lab of Experimental Immunology, NCI-FCRDC, Building 560, Room 31-93, Frederick, MD 21702-1201, USA
REMARK	Sequence update by submitter
COMMENT	On Oct 20, 1999 this sequence version replaced gi:192866.

```

1. .7036
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/sub_species="domesticus"
/db_xref="taxon:10090"
/cell_type="NK"
/tissue_type="blood"
/dev_stage="adult"
63. .4511
/note="peptidyl-prolyl cis-trans isomerase; transmembrane protein"
/codon_start=1
/product="cyclophilin-related protein"
/protein_id="AA37500.2"
/db_xref="GI:6073858"
/translation="MVTLSGGAGRAFGAVPARRRPAAALMGADRPQCHFDIEI
NRPVGRIMFQPSDICTCKNFCLSCGEGKLTGKTLCTGYSRTVVRVKNPMI
QGQDPGEQNGKGESYGYGFNDENILKHDRAFLISMANRGHTNGSOFITTPKAP
KLDGVHVPLGLVSGVEVTEQLENLTAAASRPADVVIDQGLVATLKLKDVFFKKR
KPTCEGSDSRRSSSSSESSSEVERETLRREHRRPKPRVHAHAKRKKMSEEE
PKRRTYFEGYSRSDVMEKKSVDNTRKXPVVRPEIIPVPENEFLTRDMPAIT
VPEQNI PDVAVPVDPKPSKSGSGRNI KGRGTIRVTPTRSRSHSSDQSSSESTTP
HYKEQMLRAVYPGSEKMSGKDLSDPCSSRWDRSLRSQSRSSWSYNGIYLSJSTA
RUSDGHKKHKKRKEKPKKKAKQKHCHRHQTKKRIITMPDLFSPRSPTHRMKSSC
VRRRRSSSSSHSKKDDWSKDDQDGSASTHSSDRYRSHSRSSDRSGRSSRSRA
VKSRSRLNLRSSRSRSPGPRRTSISPKPAOLSENPKVKTPLRPSVPGONGVLV
QPVAENI PVIPLSDPSPSRWKPGQKPKPSYERIOEMKAKTHLLPVQSTYSLNLI
KATVSSSSYHKKRPSDSSGSAIKYKSDASSSSGSCSKSSRSRSPSYRSPRSRS
LPTSRSLSPSSRSRSPKNSIDGSHSRSSSTYSDSDGRAMFRNKRKSGKRGK
RHRSSNPTLHSRYRGKRSRHRXYSESRSLDTSDSQSHVQVYSAPEKKGKQ
VVALNDQKGRGEGPKPEWECPFRSKENSEDHRSDSVSKNGCAGSKWSDGAKKS
VDYNTKRKSDPRRGSKBEASDSDSEVGQSHIKAKPAKPTPTPLPGSDGAWKS
RRQGSASRSESSCNLGNIRGEPOKQSKDQLKGDHTKRAKESKAKKKHKAAPK
RKAQPHQPTLPEDQDEEEMWQKQVDPKERRHYSEKCAVKDGI PVNVEKTCDSGS
SPSKRHLQBDQPLAEGHDPSCAPLKVEDNTASSPSQCHLSEHGKSGEDVLQ
TDNNMEICTPTDTPRPAKEVSPVLANHRLSDPEVNLPRQDECMHPRAGGRGERS
ESTLTGSEYKQDSSTVSPTSPTSGKKEGAEKSNLTKMKPLQGVGNLVSYSTATT
SSALDVKALSTVPEPQGLRIEIKSNHVRPGSLDFEVKTRLARLNRRPNQESSDD
SQATQDQSDGSRPHRSKQSTKTSRHRTHSVSHSRSSRSTSTSYRSYRSRSR
RDWYRGRRTSRSSSYSPSHSRSTSRSSSYDLHGRSRSYTVDSYRSRSRSR
SOWSVDHCPGSYSPSPSGVSDRSYSHRSRSPSSSKYS"

```

ORIGIN	
Alignment Scores:	
Pred. No.:	7.95e-05
Score:	241.50
Percent Similarity:	33.26%
Best Local Similarity:	20.67%
Query Match:	4.37%
DB:	10
Gaps:	60
Indels:	432
Mismatches:	428
Conservative:	162
Matches:	266
Length:	7036

US-09-721-114-2 (1-1057) x MUSCVCLOREA (1-7036)

Qy	81	llePheHisAspGlnLysValysCys	----	AspGluHisLys-AlaSerSe	95
Db	515	GIITITTCATACATCAAAAGCCTGCTCCACACCTGGATGGGTTCATGTTGTTTTGGACT	574		
Qy	89	----	----	----	----
Db	575	GGTAATATCTGTTTGTGAAGTAATTAAGACAGATTGAAATCTGAAACACAGATGCTGCAAG	634		
Qy	95	rSerProPheSerValAlaLysPheArgArgTrpAspCysSerLysCysLeuAspLysLe	115		
Db	635	CAGACCTTAT	----	----	688
Qy	115	uLysThrSerAspAsnGlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSe	135		
Db	689	G-ACAAAAGATGTTTTTGAAAAAAGGAAGAAACCACTGTCAGAAAGC	740		
Qy	135	rAspGlyCysSerIleThrPheValArgSerThrPheValProAlaSerValGlySerGl	155		
Db	741	----	----	----	754
Qy	155	nLysValSerProSerThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeu	174		
Db	755	CGGTTCCTCTCTCTCTCAGAGTCTCTCTCAGAGAGTGAAGTTGAGCGGAGACATCAG	814		
Qy	175	----	----	----	176
Db	815	AAGGACAGACATACAGAGGAGCGCCAAAGTCAGACATGCTAAAAAGAGACGGAAGAAAT	874		
Qy	177	----SerValGlnGluGlyAsnAspSerLysCysAsnAlaProSerGlyLysAsnGlyAl	195		
Db	875	GAGCAGTTCCAGAGAACCGAGAGGAAGCGCACAGTAAGCCCTCGAAGGTATCTCAGAG	934		
Qy	195	aAlaGluAlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAs	215		
Db	935	GAGTGATGTCAAT	----	----	958
Qy	215	pValAlaAlaAsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluVa	235		
Db	959	AGTTGACTCAACACATAAAGAGAAAGCGCTGTTGTCGCCCCAGAAAGACATCTCTCCAGT	1018		
Qy	235	lProGln	----	lleThrTrpHisIleGluVa	244
Db	1019	TCCGAGAACCGATTTTACTTAGAAGAGATATGCTGCTATCATCTGTGGAGCTCAACA	1078		
Qy	244	lAsn	----	----	252
Db	1079	GAACATTCAGATGTTGACCTGTTGTAAGTGATCAGAAACCCCTGTATCAAGTCTGG	1138		
Qy	253	----	----	ThrProLysLeuSerGluValVa	260
Db	1139	ACGGABAATCAAGGAAGGACGACGATTCGCTATCACACACCTCCAAGGTCAGATCCCA	1198		
Qy	260	lleuLysArgAsnGluAspGluAsnGlyLysThrGluGluThrLeuValAlaGluGlnCy	280		
Db	1199	CTCTGAGTCAAGATGATGACAGCAGTGAACCCCTCTCTACTGGAAGGAGGATGCA	1258		
Qy	280	sAsnLeuThr	----	LysAspProAsnProMetSerGlyLysGluArgAspGlnValAl	298
Db	1259	GAGACTGAGAGCCTACAGGCCCCCGAGCGAGAGAACTGGAGCAAAAGGACACAAGCTGAG	1318		
Qy	298	aGluGlnCysAsnLeuThrLysAspProLysProValSerGlyGlnLysCysGluGlnI	318		
Db	1319	TGACCCCTGTTCAGCGCATGGATGAAGACGCTGTCCAGAGATCCAGATCATGGTC	1378		
Qy	318	eCysAsnGluProCysGluGluValValLeuLysArgSerSer	332		
Db	1379	CTATATGGCATACTATTCAGATCTTAGTACAGCGAGACACTCTGATGGTCACCATAGAA	1438		
Qy	333	----	----	----	348

Db 1439 ACACAGAAAGGAAAGAGTTTAAAGCAATAAATAAGCTTAAAGAGGAGAAACATTCAG 1498  
Qy 349 -----SerLysLysArgThrAlaGlnAlaAspValSerAspAlaLysLe 363  
Db 1499 AAGACACAGACAGACAAAGAGAGAGATAGTTATCGCTGATTTGGAAACCTCA----- 1553  
Qy 363 uCysArgArgLysProLysLysValArgLeuLeuSerGluLeuIleAsnAlaGlnVal 383  
Db 1554 -----AGATCTCCCAACCCAC-----CGAATGAAGTCTCTGTGTAGAGAAAGGAGATC 1603  
Qy 383 lGluAspSerArgSerAspGluValHis-----ArgGluAsnAlaAlaAspProCysG1 401  
Db 1604 TCGTGCCT 1663  
Qy 401 uAspAspArgSerThrIle-ProValProMetGluValSerMetAsp----- 416  
Db 1664 GGATGACGGGAGTGCTTCAACCCATCTCCAGCAGAGACTCTTACAGATCCCAAGTCTCATTC 1723  
Qy 417 -----ProValSerAsnHisThrVal- 424  
Db 1724 ACGATCAGATCTTAGAGGGAGCTTAGATCAAGGGCTGTGTCAAGTCTCTCTCTCTCTCTCT 1783  
Qy 425 -----GlyGluAspGly----- 428  
Db 1784 TCTCAACAGATCAAAATCTAGATCTAGTTCAGGTCAAGGCCCGGAGAACATCAATATC 1843  
Qy 429 -----LeuLysSerSerLysAsnLysThrLysArgLysTyr 441  
Db 1844 CCCCAAAAACCTCTCAGCTGAGTGAATAATAGCCAGTT-AGACAGAACTTTTAAGGC 1902  
Qy 441 erAspValValAspAspGlySerSerLeuMet----- 451  
Db 1903 CGTCAGTGGCCACAGATGGAAATGTCTAGTGCAACCGAGTGGCAGCAGAAACATCTCTG 1962  
Qy 452 -----AsnTrpLeuAsnGlyLysLys----- 458  
Db 1963 TAATACATTGAGTGACAGCCCTCCCTCTCTAGTGGAGCCCTGGGCGAGAAACCCCTGGA 2022  
Qy 459 -----LysArgThrGlySerValHisHisThrValAlah 470  
Db 2023 AGCCCTCTTAGAGCGAATTCAGAGATGAAGCTTAAACCAACCCATCTCTCTCTCTCTCTCT 2082  
Qy 470 isProAlaGlyAsnLeuSerAsnLysLysValThrProThrAlaSerThrGlnHisAspA 490  
Db 2083 AAAGCACATCAGCTTTAAACATATTAAAGCAACCGTGAGTTCATCATCTTATTCACAAA 2142  
Qy 490 spGluAsnAspThrGluAsnGlyLeuAspThrAsnMetHisLysThrAspValCys----- 508  
Db 2143 GAGAAACCTTCAGAAAGTATGGGAGTCTTATTCAAAGTACAGTGAAGTTCTG 2202  
Qy 509 -----GlnHisValSerGluIleSerThrGlnArgCysSerSerLysGlyLysT 525  
Db 2203 GAAGCTCAGGAAGTTCGGGAGCAAGTCTCTTAGACCAAGTCTATCTCTCAGGTCTCTACA 2262  
Qy 525 hrAlaGlyLeuSerLys-----GlyL 532  
Db 2263 CAAGGTCAAGGTCAAGAAAGTCTCCCTACTTACGCTCCTCTCTAGTCTTCCATCATCTA 2322  
Qy 532 yThrHisSerAlaAlaSerThrLysLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 552  
Db 2323 GGTCCTCACTCAACAAATAAGTACAGTATGTTTCCCGACACAGTAGTCTATCTTATATA 2382  
Qy 552 leHisValLeuSerAlaGluAspGlnCysGlnMetGluThrGluAsn-----SerV 569  
Db 2383 CTTCTGTGTAGCAGTATGATGGAAAGACGACCATGTTTAGATCCAAACAGGAAAGAAAGTG 2442  
Qy 569 alLeuSerHisSer----- 573  
Db 2443 TCATCTCAATAAAGACATCGGAGCACTCTGAAAGACACTTCACAGTAAATATGTCA 2502  
Qy 574 -----AlaLysValSerProAlaGluHisAspIleGlnI 585  
Db 2503 GAGGACAGAGAAATCTCAGCTCAGAGAAAGTATAGTGAAGAGTATCATCTCTTAGATT 2562

Qy 585 leMetSerAsp-----LeuHisGluGln-----SerLeuProLysLysLysLysL 600  
Db 2563 ACATTTACAGACAGTACCAGTCCATGTTCAAGTATCTCAGCCCCAGAGAGAGAGC 2622  
Qy 600 yGlnLysLysLeuGluValThrArgGluLysGlnThrMetIleAspAspIleProMetAspI 620  
Db 2623 AGGGAAGTGGAGACATTTGAATGATGAGCAG----- 2654  
Qy 620 leValGluLeuAlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCys 640  
Db 2655 -----GGGAAAGCCAGAGAAAGGAAACCCAAAGCCTGAATGGATGTC 2700  
Qy 640 erAspIleAsnArgIleGlnSerLysThrAlaAspAspAspCysValIleValAla 660  
Db 2701 CTGTTTCTTAAAGAGAACTCCGAAGATCCTCTAGAGATGACATGTG----- 2750  
Qy 660 lAlysAspGlySerAspTyrAlaSerSerValPheAsp-----ThrAsnSerGlnGln- 677  
Db 2751 --TCCAAAGGAGAAATTCGCGGAGTAAATGGGATTCGGATCAAACTCAGAAACAG 2808  
Qy 678 -----LysSerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlyH 692  
Db 2809 ATGTGACTAAGACGAGAAAGTATCCCGGAGAGGTTTCAGAAAGGAG--GAGGGTG 2865  
Qy 692 isLeuAla-LeuThrThrGlnGlu-----SerProHisProGln 704  
Db 2866 AAGCCTTTTCAGCTCCGAGTCAGAGTGGTCAGAGTCACATCAAGCCAAACCCAG 2925  
Qy 705 AsnPheGlnSerThrGlnGlnGlnThrHisLeuArgMetGluMetValThrIle 724  
Db 2926 -----CAAAGCTCCAA-----CAGCACTTTTCTGCCCG 2955  
Qy 725 AlaAlaSerSerPro-----LeuPheSerHisAspAspGlnTyrIleAlaGluAla 742  
Db 2956 GCAGCAGCGTCTCGAAGTCTTAGAGACCACTGCTTCCAGCTCTGAGTCAGAGCT 3015  
Qy 743 ProThrGluHisTrpGly-----ArgLysAspAlaLysLysLeu 755  
Db 3016 CTTGCTCCAACTTGGGAAACATTAGAGGAGAGCCCGAGAGCAAAACACTCAAGAGATG 3075  
Qy 756 ThrTrp-----GluGlnPheLysAlaThrArg 765  
Db 3076 ATCTTAAGGGGATCACAAAAAGGCAAGAGAGAGTCAAAAGCTAAAAAGACAAA 3135  
Qy 766 AsnSerProAlaAlaThrCysGlyAlaGlnPheArgProGly-----IleGlnAla 782  
Db 3136 AACACAGGCTCCAAAC--GGAAGCAGCTTCCACTGGCACTCCACTCGAGTTTG 3192  
Qy 783 ValAspLeuThrSerThrHisValMetGlySer----- 793  
Db 3193 GTGACGATGAGGAGGAGAGATGAATGGGAAGCAAGTTACACAGGACCCAAAGAGAAA 3252  
Qy 794 -----SerSerAsnTyrAlaSerArgGlnPro 802  
Db 3253 GGCATGCTCTCAGAGTGTGAAGCTGTGAAGAGCGGCAATTCAAACGTCGAGAAACCT 3312  
Qy 803 ValIleAlaProLeuAspArgTyrAlaGluAlaValAsnGlnValHisAlaArgAsn 822  
Db 3313 GTGATG-----AAGGCAGTTTCTCCAAAGTAAACCCAAAGAG 3348  
Qy 823 PheProSerThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsnLaGly 842  
Db 3349 GTA-----CITTAGACGAGGACCCACTTGCAG 3375  
Qy 843 GlnValValLeuTyrPro-----LysGluSerMetPro 853  
Db 3376 AGGGTGGACATCATCTCCAGCTCTTGTCTCTGACCTCTGAAAGTGGAGGACAAACACGCCA 3435  
Qy 854 AlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSer-PheProAsnTyr-- 872  
Db 3436 GCTCTCCACCTAGCG-----CCAGCCTTGAAGAGCATGGCCCGGTTGAG 3483

Qy	873	----	GlyThrSerSerArgAsnGlnMetGluSerGlnLeuHis-AsnSerGlnTyAlah	891
Db	3484	GGAGGACGCTCTCAGACAGATGACAAATGAGAGATTGCACGCTGATAGG-	----	3536
Qy	891	isAsnGlnTyLysGlySerThrSerThrSerTyGlySerAsnLeuAsnGlyLysilep	911	
Db	3537	-----	ACTTCCCTCCAAAGGAGAG-	3557
Qy	911	roleuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgProLeuArgp	931	
Db	3558	-----	GTGGTGTCCTTCTAGCAACACACAGGCTAGACGCCAGAGGTGAACATTATTC	3612
Qy	931	ro-----	HisProArgValGlyValLeuGlySerLeuLeuGlnL	944
Db	3613	CAGACGAGGATGAGTGTATGCCACATCTTAGACAGCA	-----	CGAGAA-----CAAG 3660
Qy	944	ysGluileAlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyLysLeuGlyValis	964	
Db	3661	AGATAGCAGTCTGAAAGCAACCTTGGGT---	GAAGTGGGGTTAAACAGCAGCAGCT	3717
Qy	964	erThrGlyIleThrSer-----	HisGlnMetAsnArgLysGluHisPheGluAlaLeuA	982
Db	3718	CTACCACTGACAGCTCTCTAGAACCTCTGGAAGAGAGAGGGGCTGAGAGAGCC	3777	
Qy	982	snSerGlyMetPheSerAlaLysTrpAsnAlaLeuGln---	LeuGlySerValSer-----	999
Db	3778	AAATGAACCTC-----	ACAGATAAGTGAAGCATTGCAAGGTGAGGAATCTCTCAGTCT	3834
Qy	1000	-----	SerSerAlaAspPheLeuSerAlaArgAsnSerIleAlaGlnSerT	1015
Db	3835	CTACTGCAACCATCATCAGTCTCTGATGTGAGGCATTATCTACTGCGCTGAGTGA	3894	
Qy	1015	rpThrArgGly-----	LysGlyLys---MetValHisProLeuAspArgp	1029
Db	3895	AACCAACAGGCTTGAGGATAGAAATCAAAAGCAAAATAAGTTCGCGCTCTCTCT	3954	
Qy	1029	be-----	ValArgGlnAspIleCysIleThrLeuLysAsnProAlaAspPheThrThri	1047
Db	3955	TTGATGAGTAAGTAAGACGGCAGCCCTA---	AATCGAGCCCGGATCAAGAGATT	4011
Qy	1047	leSerAsnAsp 1050		
Db	4012	CCAGTGATGAT 4022		
RESULT 12				
AP520421S2				
LOCUS				
DEFINITION				
AP520422 AF441785				
ACCESSION				
AP520422.1 GI:21617579				
VERSION				
KEYWORDS				
SEGMENT				
2 of 2				
Mus musculus (house mouse)				
Mus musculus				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
1 (bases 1 to 32367)				
Desseyn, J.-L., Claverie, I. and Laine, A.				
Cloning, chromosomal localization and characterization of the				
murine mucin gene orthologous to human MUC4				
Eur. J. Biochem. 269 (13), 3150-3159 (2002)				
JOURNAL				
MEDLINE				
PUBMED				
12084055				
REFERENCE				
2 (bases 1 to 32367)				
Desseyn, J.-L., Claverie, I. and Laine, A.				
Direct Submission				
TITLE				
Submitted (11-JUN-2002) U560, INSERM, 1, place de Verdun, Lille, FR				
JOURNAL				
P-59045, France				
FEATURES				
source				
1. 32367				
/organism="Mus musculus"				
/mol_type="genomic DNA"				
/strain="129/SvJ"				

/db\_xref="taxon:10090"  
/chromosome="16"  
order(AP520421.1:<1. 1197.1. >32349)  
/gene="Muc4"  
Join(AF520421.1:<1. 178,393. 7235,7459. 7587,9257. 9390, 10152. 10316,11262. 11417,12520. 12650,14088. 14176, 15857. 16036,16239. 16355,17037. 17156,17460. 17671, 18846. 18930,19294. 19461,19907. 20008,20360. 20593, 21727. 21864,24932. 25113,25952. 26111,26974. 27147, 27843. 27916,28937. 29007,29087. 29322,30936. 31098, 31957. >32349)  
/gene="Muc4"  
/product="MUC4"  
Join(AF520421.1:106. 178,393. 7235,7459. 7587,9257. 9390, 10152. 10316,11262. 11417,12520. 12650,14088. 14176, 15857. 16036,16239. 16355,17037. 17156,17460. 17671, 18846. 18930,19294. 19461,19907. 20008,20360. 20593, 21727. 21864,24932. 25113,25952. 26111,26974. 27147, 27843. 27916,28937. 29007,29087. 29322,30936. 31098, 31957. 32152)  
/gene="Muc4"  
/note="similar to MUC4/SMC; large membrane-bound mucin"  
/codon\_start=1  
/product="MUC4"  
/protein\_id="AA066746.1"  
/db\_xref="GI:21617580"  
/translation="MRGPHRVPLCLSLCLYSCLLLPALATTQTQPMSSSSTRIT  
SQMSQASTSSDRSTKTEQTSRTDTPSSITTVSQSHSTMETSKPQTITTEV  
TTSVASSRQIQETSSQRTISPDGTTTSHAPSISASAPSTHMLTSTSTSTVS  
DSGHTALTQGLPATAQVSLTPSSNMSTVSTPITSLTQRTGKSTQSSKQVN  
IVTSLPSTSDSPAQMSVSTSSDKRTPSTGVSSTLSLTTEVLTKQTSSTDSAP  
GNTLIRIQNSTHTTKVSTTTPKLPVSTLINSQKMTLPQNGTSMEDTSGROP  
QTTTTEVTSPASSLHQIQTENSTQRTISPGRTTTSHPAFNRSSSPKTSQILTM  
PSTSTVDTKQTAITTKTPDTPDQVMSPPSKLPSTSTSTLSSLSYQLKQK  
CTSKSQTNTKNTSTPSASRDKIQTETSSQRTNSPGRKTHSGHSPSSAPSTT  
HMLSTNSQSTVDTGTTTQGTTPALITQSLTPSSNMSTVSTPITSLTQSLT  
LPOSQHTMSSTNPQTISPVVTTSPGTSQDIQTETSSQRTISPGTKTTHGAP  
NNSAPSTHMLSTSTQSTGTRHTAGTQSTPATTQSLTSLTSSPXXXXXX  
XXETETSSQRTISPGRTTSHAPMSPPSSPSTHMLSTASSTETISVDTRHTAINT  
QGSTPATTQVSPSSNMSTVSTPITSLTQSLTSSQRTISPGTKTTHGAP  
STPSASRDQIQETSSQRTISPGTKTTHGAPMSAPSTHMLSTSSQSTSGDT  
RHTAVRTQSTPATTQVSLTPSSQXXXXXXETETSLTSLTSSQSTSHASSM  
SSSPNTHLLFTTSTSTGVDTHGTVITHTGTLATTQVSLTPSSNMSTVSTP  
TSSOILTLROSQHTGKSTSNHOTTTPVVTTPSAARDQIQETSSLTSLTSPD  
CTTSHASSMSSSPNTHLLTTSSTSTGVDTHGTVITHTGTLATTXXXXXX  
XXETETSSQRTISPGTKTTHGAPMSAPSTHMLSTSSQSTSGTRHTAVRTQ  
GSTPATTQVSLTPSSQXXXXXXETETSSQRTISPGTKTTHGAPMSAPSTH  
MLSTASSTETISVDTRHTAINTQSTPATTQVSPSSNMSTVSTPITSLTSLTPK  
SQHTGKSTSNPQTTPVVTTPSSASRDQIQETSSQRTISPGTKTTHGAPMS  
SSAPSTHMLSTSTQSTGTRHTAVRTQSTPATTQVSLTPSSNMSTVSTPIT  
SPONFTLPQHTGKSTSNPSTPITPEVTITSLASRDQVQETSSQRTISPGE  
TTTSHASSMSSSPNTHLLTTSSTQSTGTRHTAVRTQSTPATTQVSLTPSSQ  
NMTVSPATSSQILTLROSQHTGKSTSNHOTTTPVVTTPSTPESATQIQET  
SLRTISPGTKTTHGAPMSAPSTHMLSTSSQSTSGTRHTAVRTQSTPATT  
TQSTPSSNMSTVSTPITSLTSLTPQHTGKSTSNPQTTPITPEMTITSPSA  
THDLETSSQRTISPGTKTTHGAPMSAPSTHMLSTSSQSTSGTRHTAVRTQ  
LMTQSTPATTQVSPSSNMSTVSTPITSLTSLTPQHTGKSTSNPQTTPITPE  
VTTPSTATTQDIQETSSQRTISPGTKTTHGAPMSAPSTHMLSTSSQSTSG  
GDTGHTAVRTQSTPATTQVSLSSONITSTVMTSTHMLSTSLTPQHTGKSTSN  
PQTTPVVTTPSTPESATQIQETSSQRTISPGTKTTHGAPMSAPSTHMLST  
ASSTETISVDTRHTAINTQSTPATTQVSLTPSSNMSTVSTPITSLTQSLT  
HTGKSTSNHOTTTPVVTTPSTPESATQIQETSSQRTISPGTKTTHGAPMS  
SPNTHLLTTSSTSTGVDTHGTVITHTGTLATTQVSLTPSSNMSTVSTPIT  
BELTSLPQHTGKSTSNPQTTPVVTTPSTSLTSSQSTSGTRHTAVRTQSTP  
TLSTPVRNTSPASTGILTSTSTGVDTHGTVITHTGTLATTQVSLTPSSNM  
TPAOLSAHSQSHQTYGTGTSOISILPDVTSKXHVASSPSTVTSFGSVSSSG  
LTTKTDNRNTAVSATSLTSPSSPSTASRTVPLSLPDQGISLFPYSGVQNL  
PARTVDNPFKILIGPLASSLDSFYVTDNGOIIPESDYDVPSPNPQFPTG  
RERVANAPFWADAFSSRGAIFQEVVTFYFNHQLIRVETLINDFTSSWGTRAK  
WTLKVTWVNPAYTAFSTFRPAMEKYRDPRLNLSKGLRGLQVLRHGH IRPRLKCL  
LMGSSGDTYFENSLTFRPAMEKYRDPRLNLSKGLRGLQVLRHGH IRPRLKCL  
RWLESPQPOQPGWGSVSCQWQOQDFRFPINPGWDRQLCSFSGRGVCCSY  
GAWGEFRGWRMHSWPQDEBQEAQWCCQNDKPSFCVWYQLRPRVSCAGYRPPR

```

AWTFGDPHITLTDNANFTNGLGDFLLVQADRNSFLLEBRTAQTGTAKATNFIAFA
AQNTSLKSPITVQWLEPSDKIRVVYNNQTVAFNTRDTEVLPIFNVTGVLITQNGS
QVSNFDTGTVTVSVIARSNILHASSLSSEYRNHTEGLLGWMDNDDPRMNGSTI
PSSNSBETLAFYGMWVNTGUGLIGIRADPLPKTEPIFLSOLLNQSAGEDLASGCK
GDRKCMEDILATGRTIGQSTNSILNEPOHMDTLAGQYPPSINCSSKIQAYKQQTWT
EITSNKDALSLSKSGQLFENGSLQWTPSPERACTLEILARDVRINLSWLQPK
IPVPCFKBQCLYNETSKGNSLEVTSCKDGTFRCLRCERPKDPCDPCFPNVNC
IPAGKRCACPNITGDGRHCAALDESPNRCFVYCYNNHGCGISEAPCQPTCTCP
PAPADNRCFLAGNSFTPTISTELPLRTIMLSLREDENASADYNASVANILENDMRA
FFNSLVELIRTSPGAQPSKSIHHWKVTSHTFYRPRGPLIHYLNNQLIGAVMEAPLL
QACQERKRSSEARKOVHPPISEADVQDOMALNLSMLREYFTCDGKYGKYLHVSQD
GUTCVSFCSGYCHNGSQCHLDPGQCSCASPTIYSSSGKHCEHLSVKLGAIFYGILP
GTUGALLGLGILAFMIFPHFCGCSXNKFYSIFDSBL"
393..7235
/ gene="Muc4"
/ note="encodes mucin-type repeats"
/ number=2
/ evidence=experimental
596..959
/ note="repeat a1"
/ evidence=experimental
1314..2344
/ note="repeats a2-a4"
/ evidence=experimental
2376..2962
/ note="repeats b1-b2"
/ evidence=experimental
2993..3559
/ note="repeats c1-c2"
/ evidence=experimental
3590..3805
/ note="repeat d1"
/ evidence=experimental
3836..6772
/ note="repeats e1-e8"
/ evidence=experimental
7459..7587
/ gene="Muc4"
/ number=3
/ evidence=experimental
9257..9390
/ gene="Muc4"
/ number=4
/ evidence=experimental
10152..10316
/ gene="Muc4"
/ number=5
/ evidence=experimental
11262..11417
/ gene="Muc4"
/ number=6
/ evidence=experimental
12520..12650
/ gene="Muc4"
/ number=7
/ evidence=experimental
14088..14176
/ gene="Muc4"
/ number=8
/ evidence=experimental
15857..16036
/ gene="Muc4"
/ number=9
/ evidence=experimental
16239..16355
/ gene="Muc4"
/ number=10
/ evidence=experimental
17037..17156
/ gene="Muc4"
/ number=11
/ evidence=experimental
17460..17671

```

```

/ gene="Muc4"
/ number=12
/ evidence=experimental
18846..18930
/ gene="Muc4"
/ number=13
/ evidence=experimental
19294..19461
/ gene="Muc4"
/ number=14
/ evidence=experimental
19907..20008
/ gene="Muc4"
/ number=15
/ evidence=experimental
20360..20593
/ gene="Muc4"
/ number=16
/ evidence=experimental
21727..21864
/ gene="Muc4"
/ number=17
/ evidence=experimental
24932..25113
/ gene="Muc4"
/ number=18
/ evidence=experimental

Alignment Scores:
Pred. No.: 0.00141 Length: 32367
Score: 234.50 Matches: 195
Percent Similarity: 34.22% Conservative: 138
Best Local Similarity: 20.04% Mismatches: 399
Query Match: 4.24% Indels: 242
DB: 10 Gaps: 42

US-09-721-114-2 (1-1057) x AF520421S2 (1-32367)

Qy 60 SerIleArgGlyTyr-----ValAlaLeuLeuGlnLysLysAspProLysPheCys 76
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 319 TCAGTTAGAGTTACCATTTATCTAAACAATCTGCAAGATAATAAAGTCGTTGTTGT 378

Qy 77 SerLeuSerArgIlePheHisAspGlnLysLysCysAspGluHisLys-AlaSerSerSe 96
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 379 TTATATTCAAGA-----AGATGCTCTAGCAACAACATCAACAGAC 420

Qy 96 rProPheSerValAlaLysPheArgTrpAspCysSerLysCysLeuAspLysLeuLy 116
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 421 ACCATGTCCTTCCAGCAGCACCGA-----ACATCCCAATGAGTCCACAGCATC 474

Qy 116 sThrSerAspAsnGlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAs 136
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 475 GACCTCATCAACATCTTCTGACAGAAGGACAAAGTAAACAGAACATCAACAGGGA 534

Qy 136 p---GlyCysSerIleThrPheVal-----ArgSerThrPh 147
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 535 TACTCCAGTTCCATAAACACAGTTTCTCAGAGCCATCATACTACAGATGGAGACCAG 594

Qy 147 eValProAlaSerValGlySerGlnLysValSerProSerThrGlnSerSerGlnGlyLy 167
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 595 CAACCTCAACACCACCAACCCAGGAGTGCACCATCACTCCTTCAGCTCCTCAGC 654

Qy 167 sAenAlaAspArgSerThrLeuProLysSerValGlnGluGlyAsnAspSerLysCysAs 187
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 655 TGACCAAAATACAGACA-----GAGCAAGATTCTCAAGAAGACAAT 693

Qy 187 nAlaProSerGlyLysAsnGlyAlaAlaGluAlaAsnThrAspSerProMetLysAspLe 207
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 694 CTCTCTCTGATGGA-----ACAACCCTTCACATGCTCCCGAGTATCAGCAG 738

Qy 207 uGlnGlyProAlaGlnAsnTyrAspValAlaAlaAsnValSerGluAspAsnThrSerVa 227
|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 739 CTCAGCTCCCAAGTACAAACACATGTTAAACCAACATCTCTCCACAGAAAGTACTCAGT 798

```



```

*-----*
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

```



misc_feature	185455..188044	/note="assembly_fragment:00206.0"			
misc_feature	188145..191198	/note="assembly_fragment:00247"			
misc_feature	191299..227282	/note="assembly_fragment:01877"			
ORIGIN					
Alignment Scores:					
Pred. No.:	0.016	Length:	227282		
Score:	234.50	Matches:	213		
Percent Similarity:	33.92%	Conservative:	171		
Best Local Similarity:	18.82%	Mismatches:	439		
Query Match:	4.24%	Indels:	310		
DB:	2	Gaps:	40		
US-09-721-114-2 (1-1057) x BX842684 (1-227282)					
Qy	76	CysSerLeuSerArgIlePheHisAspGlnLysLysCysAspGluHisLysAlaSerSer	95		
Db	131791	TGCTCGCTGTCAGGAATATGACATCTGGAAGATCTCTGACCTCCAGGAGCTGT	131732		
Qy	96	SerProPheSerValAlaLysPheArgArgTirAspCysSerLysCysLeuAspLysLeu	115		
Db	131731	AAGCGCATCAGATCTCAAGCTC-----AAGTGCATCCTCAAGAG	131690		
Qy	116	LysThrSerAspAsnGlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSer	135		
Db	131689	AATAGTGTGCAAAAGGATCTACTGACGACATTCGCACAGTGTCCAGTGAGGAGTACAC	131630		
Qy	136	AspGlyCysSer-----lleThrPheValArgSerThrPheValProAlaSerValG1	153		
Db	131629	AGACGATGTAGTGAAAGGCACTTGCATTCAACAGACCTTC-----	131587		
Qy	153	YserGlnLysValSerProSerThrGlnSer-----SerGlnG1	166		
Db	131586	-GGACAGCAAGTAGATACACAAATTTGAACAGCTGTACCATCAACATTTGCTGAGTCGAG	131528		
Qy	166	YLeuAsnAlaAspArgSerThrLeuProLysSerValGlnGluGlyAsnAspSerLysCy	186		
Db	131527	TGAAGTATCTTCATATCCCTTAACCAAGAGCAGCAAGTACTGTAGATGATGAGTGTA	131468		
Qy	186	sAsnAlaProSerGlyLysAsnGlyAlaAlaGluAlaAsnThrAspSerProMetLysAs	206		
Db	131467	GAACTTCATGCTCATAT-----GAAGAGACAGACCTGTATGATTTCTCCAA	131414		
Qy	206	pLeuGlnGlyProAlaGlnAsnTyrAspValAlaAlaAsnValSerGluAspAsnThrSe	226		
Db	131413	CCTCCCTAAAGATCATTTAAGTGTTCAGATCCTCAGCAAGTAGGAGGAGGATGAGGTC	131354		
Qy	226	rValAspValGlyAlaLeuProGluValProGlnIleThrTriPheHisIleGluValAsnG1	246		
Db	131353	TATATCTGTC-----TTGAGCTCTTCTCACAATTTCTAGCATCCTGAGGAGGATCA	131300		
Qy	246	yAlaAspGlnProProSerThrProLysLeuSerGluValValLeuLysArgAsnGluAs	266		
Db	131299	AGATGATGAC-----GTCACCCCATCTCTCGAGAGCATCCCATGACCAACAGGATGA	131246		
Qy	266	pGluAsnGlyLysThrGluGlnThrLeuValAlaGluGlnCysAsnLeuThrLysAspPr	286		
Db	131245	TGAGGAGGCAAGATCTCTGAAACA-----AACCAATGTTCTCTCCCTCCCTGTAA	131195		
Qy	286	oAsnProMetSerGlyLysGluArgAspGlnValAlaGluClnCysAsnLeuThrLysAs	306		
Db	131194	AGTCTCTGATCTCTGAGGCAT-----CTGTC	131168		
Qy	306	pProLysProValSerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluVa	326		
Db	131167	ACCAAGACCACCGAGC-----	131152		
Qy	326	lValLeuLysArgSerSerLysSerLysArgLysThrAspLysLysLeuMetLysLysG1	346		
Db	131151	-----AAGCTTCATCTTCAGGCGGTCTAAAACTTACAGAAAGTTTCCGACCTCTGA	131099		

Qy	346	nGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAspAlaLysLysLeuCyArgar	366		
Db	131098	AGTTGGCACTGAAGAGTCTGAGCACTCTCTCCCAACATTTTGTCTGGACC	131039		
Qy	366	glysProLysLysValArgLeuLeuSerGluIleleAsnAlaAsnGlnValGluAspSe	386		
Db	131038	CAAAACCTGTAAATGTGTAGTCACAAAGGACACTCCAGAAGCTTCGAACCAATGATGA	130979		
Qy	386	rArgSerAsp-----GluValHisArgGluAsnAla-----	396		
Db	130978	GGTAGAAGACATGCAAGTTTCACTAAGTCAAGGAGGCACTAAGTGCACCTCTACAAA	130919		
Qy	397	-----AlaAspProCysGluAspAspArgSerThrIleProValProMe	411		
Db	130918	GTCTGTCTGAGTTATCAACCTCTGTGAGAAATCGAAATCTCTGGAATATATACAGA	130859		
Qy	411	tGluValSerMet-----AspIleProVa	419		
Db	130858	GAAGCATCATTAAGTGTCTCTCTATAAAATCCCAACTCTAGAAAATCAGTTAAAC	130799		
Qy	419	lSerAsnHisThrValGlyGluAspGlyLeuLysSerSerLysThrLysArgly	439		
Db	130798	AAATGTTTGTAGATTACGATGAAGAGCACTAAGTGCACATCAACCAAGTCTCAAGAGT	130739		
Qy	439	sTyrSerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysLy	459		
Db	130738	CTCAGGTGATGTGTGAGAGGAAAC-----CTAGAAGTAGAAAATC	130694		
Qy	459	sArgThrGlySerValHisHisThrValAla-----HisProAl	472		
Db	130693	AAAGTCCATGTCTGTCACTCAATATATCATGAGATCCAGAGTATCAGATCATGGTAT	130634		
Qy	472	agLysAsnLeuSerAsnLysLysValThrProThrAlaSerThrGlnHisAspAspGluAs	492		
Db	130633	TGCTGAGCTCGGCAAA-----	130618		
Qy	492	nAspThrGluAsnGlyLeuAspThrAsnMetHisLys-----ThrAspValCysGlnHi	510		
Db	130617	-----TTACAAGACATGTTTCAGAAAGGGCTGTAAAGTGTGGCTCATC	130574		
Qy	510	sValSerGluIleSerThrGlnArgCysSerSerLysGlyLysThrAlaGlyLeuSerLy	530		
Db	130573	TAGGTCAATATTTCCACTCAT-----GGGGCAATTTGAGGA	130538		
Qy	530	sGlyLysThrHisSerAlaAlaSerThrLysTyrGlyGlyGluSerThrArgAsnGlyG1	550		
Db	130537	GGGCATAGAAGAAAAGCTTTGCTGCTGAATCTGA-----ACATCAACAAATC	130487		
Qy	550	nAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrGluAsn-----Se	568		
Db	130486	TGGAATTGAAGAACTAAACATGACGATCCACAGAGAAAGGGCTTCCCGCTCTGTTTC	130427		
Qy	568	rValLeuSerHisSerAlaLysValSerProAlaGluHisAspIleGlnIleMetSerAs	588		
Db	130426	GATAATATCAATATTTTTCGAAGATCTAAACCTCTTTGAACCTGAGTGAGTACGAGTA	130367		
Qy	588	pLeuHisGluGlnSerLeuPro-----LysLy	597		
Db	130366	TTACAATCAAGACCGAGCCCGAGCTGTTTCCCTCAATCAATCAATGTCTCAACAAAAC	130307		
Qy	597	sLysLysLysGlnLysLeuGluValThr---ArgGluLysGlnThrMetIleAspAspI1	616		
Db	130306	AAATCTGAATGTCTCAGAGAAATTTCAATCCATGAAGAGGAGGAATCTTTTAATTCAT	130247		
Qy	616	e-----ProMetAspIleValGluLeuLeuAlaLysAsnG1	628		
Db	130246	CAACACTCTCTGTTGAATTAACAAACCTGAAGAT-----AAAGGAGC	130205		
Qy	628	nHisGluArgGlnLeuMetThrGlu-----ThrAspCy	639		
Db	130204	TCATCAAGTGGAGTACATGTGGAGAGATCTCCAGCTCTGCATCATCAAAATCAATAT	130145		

Qy 639 sserAspIleasnArgIleGlnSerLysThrThrAlaaspAspCysValIleValAl 659  
Db 130144 CTCTACATATCCAGAGCATCTCAATTTTGAAGTAAACCAAGAC---GAATATCTGGA 130088  
Qy 659 aAlaLysAspGlySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGlnLysSe 679  
Db 130087 AAAAGAGGATGATCAGGAAGAGCATCAAGTGGCAATTCGTGTAAGTCT---TC 130037  
Qy 679 rLeuAlaSerGlnSerThrGlnLysGlu----- 688  
Db 130036 TGTATCTGCAGATCAAGAAATCTGAGACACACCAATATCAAGATCAGAGACAAAT 129977  
Qy 689 -----LeuGlnGlyHisLeuAlaLeuThrThrGlnG 699  
Db 129976 CAAAGATCGGCTCCAAAGTGTATGCTCGCTCAATCAATCAATTCATTGACATCAATAA 129917  
Qy 699 uSerProHisProGlnAsnPheGlnSerThrGlnGlnGlnThrHisLeuArgMetG 719  
Db 129916 GTCC-----AAGATTGGAATGAGAAACAGAAACAGAGCTTC 129878  
Qy 719 uGluMetValThrIleAlaAlaSerSerProLeuPheSerHisHisAspGlnTyrI 739  
Db 129877 CAGTGCTCTGTCTTAATCGGAAGCTTCACATTGTCA----- 129838  
Qy 739 eAlaGluAlaProThrGluHisIstPpGlyArgLysAspAlaLysLysLeuThrTrpGluG 759  
Db 129837 -----AAAAATTAGACATTGAACA 129818  
Qy 759 nPheLysAlaThrThrArgAsnSerProAlaAlaThrCysGlyAlaGlnPheArgProG 779  
Db 129817 ACTGGAAGAAAGACATCAAGTGCATTCATCGCTTAATCAATAGTCTTAAAGATCTAA 129758  
Qy 779 yIleGlnAlaValAspLeuThrSerThrHisValMetGlySerSerSerAsnTyrAlaSe 799  
Db 129757 AAGTCCAAATGTCCAGAGGTAGGAGGATGATGATGATCATCTGCTGTTCATGCAGC 129698  
Qy 799 rArgGln-ProValIleAlaProLeuAspArgTyrAlaGluArgAlaValAsn----- 816  
Db 129697 ATCAGGTACTCTCTCAACAGCCCTTAAGAAACATGTGAAGTTCAGATCAATACCCAGA 129638  
Qy 817 -----GlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGluAlas 833  
Db 129637 GGAAAGAGGCGCAAGTGTCTCAACAAATCCA-----A 129602  
Qy 833 erLysLeuCysAspArgArgAsnAla-----GlyGlnValValLeuTyrP 848  
Db 129601 TACCTCTGTTAAATCAAGAATAATCAATGTATCTTAATATGACAGCATAGTAAATGGA 129542  
Qy 848 rLysGluSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlas 868  
Db 129541 TAAAAATAGATGATTTGAGATCTGCTAGTGCATCT-----CCTTAATCTGCAT 129488  
Qy 868 erPheProAsnTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerG 888  
Db 129487 ATCTGCAAAATCAGAAATTCATCAGAGAAGAT-CAGGTAGAACAATTCAGCATGCAGAA 129429  
Qy 888 lntYrAlaHisAsn----- 892  
Db 129428 AACGAGCCAGAGGCTTTGTGACAGCGTCTAACCAATCAGGAGTAATGTAGTGAAG 129369  
Qy 893 -----GlnTyrLysGlySerThrSerThrSerThrSerTyrGlySerAsnLeuAsnG 908  
Db 129368 AAGTTGAAGTTGCAGAGCAAGCGCTCTGAGTGCATCTCAATCAATCAATCAATGCTCTG 129309  
Qy 908 ly-----LysIleProLeuThrPheGluAspLeuSerArgH 920  
Db 129308 GAAGATCAATATAGTCCAAACTTTAAAGTGCCCACTGCTCACTCAAGAGCAGAAC 129249  
Qy 920 isGlnLeuHisAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlys 940  
Db 129248 GCCAACTAGATGACGAGCTCAG-----AGTG 129222  
Qy 940 erLeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrL 960

Db 129221 CACTTTGGAAAATCAATCAAGCATT-----TCAGCAAGATCCAGA 129183  
Qy 960 yLeuGlyValSerThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluA 980  
Db 129182 AGTCTAAATCTTTCAGATGTTTTCAGATGATGAGCTTCTGATGCAAAATACCAAGAAGAGG 129123  
Qy 980 lAlaLeuAsnSerGlyMetPheSerAlaLysTrpAsn-----AlaLeuGlnLeuGlySerV 998  
Db 129122 AAAGACTTCAAGTCTGTTCTGCAAGATCAACATTTCTGCAATATCAATGAAGTCAA 129063  
Qy 998 alSerSerAlaAspPhe----- 1004  
Db 129062 ATGCTTCTATGACAGATGTCAACACTGATGACAATGAACAGGGGACAGAGTAGCAAGTG 129003  
Qy 1005 -----LeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgG 1018  
Db 129002 CAGCATCAGTAACATCTGATATTTCTACCAAAACCAAGTATCTCTGCAAAATCCAGGCAT 128943  
Qy 1018 lLysGlyLysMetValHisProLeuAspArgPheValArgGlnAspIleCysIleThrA 1038  
Db 128942 CCAAGTCTTTCAGAGTCCGACCCGAGAGCGCTGATGATAAAGAAGATTTATGAACATAG 128883  
Qy 1038 snLysAsnProAlaAspPheThrThrIleSer 1048  
Db 128882 ATAGGTCTCCAGTGCATTTTCAGTAATATCT 128851  
RESULT 14  
AC009256/c 184271 bp DNA linear INV 13-DEC-2001  
LOCUS Drosophila melanogaster, chromosome X, region 14A-14B, BAC clone  
DEFINITION BACR03G18, complete sequence.  
ACCESSION AC009256  
VERSION AC009256.8 GI:17646883  
KEYWORDS HTG.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 184271)  
AUTHORS Celisner,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,  
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Branden,R.C.,  
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,  
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,  
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,  
Fierfies,S., Frise,B., Galle,R.F., Garg,N.S., George,R.A.,  
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,  
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,  
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunco,J.,  
Paclet,J., Paragay,V., Park,S., Patel,S., Pfeiffer,B., Pflueger,  
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,  
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,  
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.  
Sequencing of Drosophila chromosome X, region 14A-14B  
Unpublished (1998)  
TITLE 2 (bases 1 to 184271)  
JOURNAL  
REFERENCE  
AUTHORS Celisner,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,  
Buthenoff,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L.,  
Doyle,C.M., Farfan,D.E., Galle,R.A., George,R.A., Harris,N.L.,  
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,  
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,  
Moshrefi,A., Moshrefi,M., Nixon,K., Paclet,J.M., Park,S.,  
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.  
TITLE Direct Submission  
JOURNAL Submitted (09-AUG-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
COMMENT On Dec 13, 2001 this sequence version replaced gi:5978361.  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

## FEATURES

source

```

1. 184271
   /organism="Drosophila melanogaster"
   /mol_type="genomic DNA"
   /strain="y; cn bw sp"
   /db_xref="taxon:7227"
   /chromosome="X"
   /map="14A-14B"
   /clone="BACR03G18 (D892)"
   /clone_lib="RP11-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial 3c0r1 in
pBACE3.6)"

```

## ORIGIN

```

Alignment Scores:
Pred. No.:      0.0152      Length:      184271
Score:          233.00      Matches:      241
Percent Similarity: 36.00%      Conservative: 196
Best Local Similarity: 19.85%      Mismatches:   503
Query Match:      4.22%      Indels:       280
DB:               3         Gaps:         48

```

US-09-721-114-2 (1-1057) x AC009256 (1-184271)

```

QY 25 ThrGlyAlaValAlaProValLeuGluLeuThrAlaThrProArgGlnAspAlaAla 44
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 171046 ACATTTCGATTCTCAATCAATAGCAATAGCAATAGTCCATTACAGATTGAGTGCA 170987
QY 45 GluAlaGlyValAspGluProAlaGlnHisGlnCysGluHisPheSerIleArgGlyTyr 64
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170986 GAA-----GATCAACAATCAAGAGAAACATCAGAA----- 170954
QY 65 ValAlaLeuLeuGlnLysLysAspProLysPheCysSerLeuSerArgIlePheHisAsp 84
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170953 -----CTTTCAAAATCTTTACCTCAATTAAACACAGAGAAAGCATTTTCATTCAA 170903
QY 85 GlnLysLysCysAspGluHisLysAlaSerSerSerProPheSerValAlaLysPheArg 104
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170902 GAATCTAGTCGACAAACAATCAATGACAGAGACCG----- 170864
QY 105 ArgTrpAspCysSerLysCysLeuAspLysLeuLysThrSerAspAsnGlyThrAlaPro 124
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170863 ---TGGACACTATCAACATCTCTCTCAACAAGTTCAAGACAAACAATACATTT--- 170810
QY 125 ArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIleThrPheValArg 144
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170809 -----TCGAGTCAGGTGAATGAGACAAACTTCTCTAGAGGTACAGG 170762
QY 145 SerThrPheValProAlaSerValGlySerGlnLysValSerProSerThr----- 161
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170761 ACATGTCGATTCTCAATCAATT---ACAATGACAAATAGTCCACTTACAGATTTGAGT 170705
QY 162 -----GlnSerSerGlnGlyLysAsnAlaAspArgSer---ThrLeuProLysSer 177
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170704 GCAGAGATCAACAACAATCAAGAGAAACATCAGACTTTCAAAATCTTTACCTCAATTA 170645
QY 178 ValGlnGlyLysAsn-----AspSerLysCysAsnAlaProSerGlyLysAsn 193
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170644 ACAACAGAGAAGACATTTTCATTTCAGAAATCTAGTGCAGACAAACAATCAATGACAGA 170585
QY 194 GlyAlaAlaGluAlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsn 213
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170584 AGAGCCGTGCACACTATCAACATCTCTTC-CTCAACAAGTCTCAAGACGAAACATTACAT 170526
QY 214 TyrAspValAlaAlaAsnValSerGluAspAsnThrSerVal---AspValGlyAlaLeu 232
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

DB 170525 TTTCGAGTCAGAGTGTAAATAT---GAAGACAAAACCTTCTTAGAGGATACAAAGGACACTG 170469
QY 233 ProGluValProGlnIleThrTrpHis-----IleGluValAsnGlyAlaAsp 248
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170468 TCGATTCTTCATCAATCAATAGCAATAGTCCACTTACAGATTGAGTGCAGAGAT 170409
QY 249 Gln-----ProProSerThrProLysLeuSerGluValValLeuLysArgAsnGluAsp 266
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170408 CAACAACAATCAAGAGGAAACATCAGAACTTTCAAAATCTTTACCTCAATTAACAACAGAA 170349
QY 267 GluAsnGlyLysThrGluGluThrLeuValAlaGlnCysAsnLeuThrLysAspPro 286
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170348 GAAGCATTTTCATTTCAGAAATCT---AGTCAGAACAAATCAAAATCAGACAGAGCGG 170292
QY 287 AsnProMetSerGlyLysGluArgAspGlnValAlaGlnCysAsnLeuThr----- 304
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170291 TGGACACTATCA---ACATCTCTCTCAACAAGTTCAAAGACAAACATTAATTTTCG 170235
QY 305 -----LysAspProLysProValSerGlyGln 313
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170234 AGTCAGAGTGTATATGAAGACAAAACCTTCTTAGAGGATACAAAGACACTGTGATTCT 170175
QY 314 LysCysGluGlnIleCysAsnGluProCysGluGluValValLeuLysArgSerLys 333
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170174 CAATCAATTAATGACAAATAGTCCACTTACAGATTGAG-TGCAGAGATCAACAA 170116
QY 334 SerLysArg-----LysThrAspLysLysLeuMetLysLysGlnGlnHis 348
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170115 TCAAGAGGAAACATCAGAACTTTCAAAATCTTTACCTCAATTAACAACAGAGAAAGCAT 170056
QY 349 -----SerLysLysArgThrAlaGlnAlaAspVal 358
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170055 TTCATTTCAGAAATCTAGTGCAGACAAATCAATGACAGAGCGGTGCACACTATC 169996
QY 359 SerAspAlaLysLeuCysArgLysProLysValArgLeuLeuSerGluLeuIle 378
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169995 AACATCTCTCTCAACAAGTTCAAAGACAAAC--ATTACATTTTCGAGTCAGAGTGT 169938
QY 379 AsnAlaAsnGln-----ValGluAspSerArg-----SerAspGluValHis 392
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169937 AATGAAGACAAACCTTCTTAGAGGATACAGGACACTGTGCTTCTCAATCAATTACA 169878
QY 393 ArgGluAsnAla-----AlaAspProCysGluAspAspArgSerThrIleProValPro 410
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169877 ATGACAAATAGTCCACTTACAGATTGAGTCAGAAAGATCAACAATAACAGAGGAAACA 169818
QY 411 MetGluValSerMetAspIlePro----- 418
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169817 TCAGAACTTTCAAAATCTTTACCTCAATTAACAAGAGAAAGCATTTTCATTCAGAA 169758
QY 419 -----ValSerAsnHisThrValGlyLysAsp-----GlyLeuLys 430
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169757 TCTAGTGCAGACAAATCAATGACAGAGCGGTGCACACTATCAACATCTCTCTCT 169698
QY 431 SerSerLysAsnLysThrLysArgLysTyrSer-----AspValValAspGlySer 448
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169697 CAACAAGTTCAAAGACAAACAATTCATTTTCGAGTCAGAGTGTAAATGAGACAAAAC 169638
QY 449 SerLeuMetAsnTrpLeuAsnGlyLysLysArgThrGlySerValHisThrVal 468
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169637 TCTCTA-----GAGGATCAAGACACTCTCGATTCTCAATCAATA 169596
QY 469 AlaHisProAlaGlyAsnLeuSerAsnLysValThrProThrAlaSerThrGlnHis 488
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169595 -----ACAATGACAAATAGTCCACTTACAGATTGAGTGCAGAGAGATCAA 169551
QY 489 AspAspGluAsnAsp-----ThrGluAsn 496
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169550 ACAATCAAGAGGAAACATCAGAACTTTCAAAATCTTTACCTCAATTAACAACGGAAGA 169491
QY 497 GlyLeu-----AspThrAsnMetHisLysThrAspValCysGlnHisValSerGlu 513
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 169490 AGCATTTTCATTCAAGAAATCTAGTGCAGAAACAATAATCAATCAACAGAGCGGTGACA 169431

```

Qy 514 IleSerThrGlnArgCysSerSerLysGlyLysThrAlaGlyLeuSerLysGlyLysThr 533  
 Db 169430 CTATCAACGCTCTTTCTCAACCAAGTTCAAGACCAACATATATATTTTCGAGTCAGAGT 169371  
 Qy 534 HisSerAlaAlaSerThrLysThrLysGlyLysThrArgAsn---GlyGlnAsnIle 552  
 Db 169370 GTAATGAAGACAAACATCTCTAGGAATCAAGGACACTGTCGATTTCTCAATCAATA 169311  
 Qy 553 HisVal-----LeuSerAlaGluAspGlnCysGlnMetGluThr 565  
 Db 169310 ACAATGACAAATAGTCCACTTACAGATTTGAGTCGAGAGATCAACAAATCAAGAG--- 169254  
 Qy 566 GluAsnSerValLeuSerHisSerAlaLysValSerProAlaGluHisAspIleGlnIle 585  
 Db 169253 GAAGCATCAGAACATTTCAATCTTACCTCAATTAACAACAGAGAAAGCATTTTCATTT 169194  
 Qy 586 MetSerAspLeuHisGluGln-----Ser 593  
 Db 169193 CAAGAATCTAGTCGAGACCAAAATCAAAATGACAGAGCGCGGACACATCTATCAACATCT 169134  
 Qy 594 LeuProLysLysLysLysGlnLysLeu-----GluValThrArgGlu 608  
 Db 169133 CTTCCTCAACACAGTTCGAAGACAAACATATATTTTCGAGTCAGAGTGTAAATGAAGAC 169074  
 Qy 609 LysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAla----- 625  
 Db 169073 AAAAATTTCTCTAGGAGATACAGGACACTGTCGATTTCTCAATCAATAACAATAACAAT 169014  
 Qy 626 -----LysAsnGlnHisGluArg----- 631  
 Db 169013 AGTCTCTTACAGATTTGAGTCGAGAGATCAACAAATCAAGAGGAACATCAAGACATTT 168954  
 Qy 632 -----GlnLeuMetThrGluThrAsp----- 638  
 Db 168953 TCAAAATCTTACCTCAATTAACAACAGAGAAAGCATTTCAATTCAGAGATCTAGTGCA 168894  
 Qy 639 -----Cys 639  
 Db 168893 GAACAAATCAAAATGACAGAGAGCGGTGGACATCAACATCTTCTCCTCAACAAAGT 168834  
 Qy 640 SerAspIleAsnArgIleGlnSerLysThrAlaAspAspAspCysValIleValAla 659  
 Db 168833 TCAAGACAAACATATATTTTCGAGTCAGAGTGTAAATGAAGACAAACATTTCTCTAGAG 168774  
 Qy 660 AlaLysAspGlySerAspThrAlaSerSerValPheAspThrAsnSerGlnGlnLysSer 679  
 Db 168773 GATCAAGGACACTGTCGATTTCTCAATCAATCAATCAATCAATCAATCAATCAATCAAT 168714  
 Qy 680 Leu---AlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAla----- 694  
 Db 168713 TTGAGTCGAGAGATCAACAAATCAAGAGGAACATCAGAACTTTCAAAATCTTTACCT 168654  
 Qy 695 ---LeuThrThrGlnGluSerProHisProGlnAsnPheGlnSerThrGlnGlnGln 713  
 Db 168653 CAATTAACAACAGAGAAAGCATTTCAATTTCAAGAAATCTAGTCGAGAAACAAATCAAT 168594  
 Qy 714 ThrHisLeuArgMetGluGluMetValThrIleAlaSerSerProLeuPheSerHis 733  
 Db 168593 ACA-----GAAGAGCGGTGGACACTATCAACATCTCTCTCTCAATCAAGTTC- 168547  
 Qy 734 HisAspAspGlnThrIleAlaGluAlaPro-----ThrGluHisTrpGlyArgLys 750  
 Db 168546 AAAGACAAACAAATATATTTTCGAGTCAGAGTGAATGAAGACAAACATCTCTAGG-- 168489  
 Qy 751 AspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSerProAlaAla 770  
 Db 168488 GATCAAGGACACTGTCGATTTCTCAATCAAGAGAAATCAAAATAGT----- 168441  
 Qy 771 ThrCysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisVal 790  
 Db 168440 -----CTACTTACAGATTTGAGTCGAGAGATCAACAAATCAAGAG--- 168399

791 MetGlySerSerSerAsnTyralaSerArgGlnProValIleAlaProLeuAsp----- 808  
 Db 168398 -----GAACATCAGAACTTTCAAAATCTTTACCTCAATTAACAACAGAGAAAGCATTT 168345  
 Qy 809 ArgTyralaGluArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAla 828  
 Db 168344 TCATTTCAAGAAATCTAGTCGAGAAACAAATCAAAATCAAAATGACAGAGCGGTGGACACTATCA 168285  
 Qy 829 ThrMet-----GluAlaSerLysLeuCysAspArgAsnAlaGlyGlnValValLeu 846  
 Db 168284 ACATCTCTTCTCAACACAGTTCAAAGACAAACATATACATTTTCGAGTCAGAGTGTATAT 168225  
 Qy 847 Ty-ProLysGluSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeu 866  
 Db 168224 GAAGACAAACATTTCTCTAGAGATACAGGACACTGTCGATTTCTCAATCAATCAATCAATG 168165  
 Qy 867 AlaSerPheProAsnTyrglyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsn 886  
 Db 168164 ACAATAGTCTACATTTAGATTTGAGTCGAGAAATCAAAATCAAAATCAAAATCAAAAT 168123  
 Qy 887 SerGlnTyralaHisAsnGlnTyrglySerThrSerThrSerTyrglySerAsnLeu 906  
 Db 168122 -----AATCAAGAGAAACATCAGAACTTTCAAAATCTTTACCTCAATTA 168078  
 Qy 907 AsnGlyLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHis 926  
 Db 168077 ACAACAGAGAAAGCATTTCAATTTCAAGAAATCTAGTCGAGAAACAAATCAAAATCAAGAA 168018  
 Qy 927 ArgProLeuArgProHisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIle 946  
 Db 168017 AACCGTGGACACCATCAACATCTCTTCTCAACAAAGTTCAAAGACAAACATCAATCATTT 167958  
 Qy 947 AlaAsnTrpSer-----GluAsnCysGlyThrGlnSerGlyTyrglyLysLeuGlyValSer 964  
 Db 167957 TCGAGTCAGAGTGTATATGAAGACAA-AACTTCTCTAGAGATACAGGACACTGTCGAT 167899  
 Qy 965 ThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGly 984  
 Db 167898 TTCTCAATCAAGAGAAATCAAAATAGTCTACTTACAGATTTGAGTCGAGAGATCAAAAC 167839  
 Qy 985 MetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAspPhe 1004  
 Db 167838 AAA-----TCAAGAGGAAACATCAGAACTTTCAAAATCTTT 167803  
 Qy 1005 LeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHis 1024  
 Db 167802 ACCTCAATTAACAACAGC----- 167785  
 Qy 1025 ProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAsn-----Pro 1041  
 Db 167784 ---AGAAAGCATTTCAATTCAGAAATCTAGTCGAGAAACAAATCAATCAAGAGAGGCC 167728  
 Qy 1042 AlaAspPheThrThrIleSerAsnAspAsnGluTyrglyMetAsp 1055  
 Db 167727 GTGACACTATCAACATCTCTCTCAACAAAGTTCAAAGAC 167686

RESULT 15  
 AC010919  
 LOCUS  
 DEFINITION  
 Drosophila melanogaster, chromosome X, region 14A-14A, BAC clone  
 BACR32K23, complete sequence.  
 AC010919  
 AC010919  
 VERSION  
 AC010919.10 GI:17646882  
 KEYWORDS  
 HTG.  
 SOURCE  
 Drosophila melanogaster (fruit fly)  
 ORGANISM  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 198961)  
 REFERENCE  
 Celiker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,  
 Holt, R.A., Evans, J.D., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,  
 Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,

Carlson, J.W., Center, A., Champagne, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferrera, S., Frise, E., Galle, R.F., Gaig, N.S., George, R.A., Gonzalez, M., Hoack, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, T.C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McInosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome X, region 14A-14A Unpublished (1998)

2 (bases 1 to 198961)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

Direct Submission

Submitted (27-SEP-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Dec 13, 2001 this sequence version replaced gi:5563419.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

Location/Qualifiers

1..198961

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/strain="y; cn bw sp"

/db\_xref="taxon:7227"

/chromosome="X"

/map="14A-14A"

/clone="BACR32K23 (D891)"

/clone\_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"

Alignment Scores:

Pred. No.:	0.0167	Length:	198961
Score:	233.00	Matches:	241
Percent Similarity:	36.00%	Conservative:	196
Best Local Similarity:	19.85%	Mismatches:	503
Query Match:	4.22%	Indels:	280
DB:	3	Gaps:	48

US-09-721-114-2 (1-1057) x AC010919 (1-198961)

Qy 25 ThrGlyAlaValAlaProValLeuLeuThraThraThraProGlnAspAlaAla 44

Db 179335 ACATTTTCGATTTCTCAATCAATAGCAATGACAAATAGTCCATTACAGATTGAGTGA 179394

Qy 45 GluAlaGlyValAspProAlaGlnHisGlnCysGluHisPheSerIleArgGlyTyr 64

Db 179395 GAA-----GATCAACAATCAAGAGAAACATCAGAA-----179427

Qy 65 ValAlaLeuLeuGlnHisLysAspProLysPheCysSerLeuSerArgIlePheHisAsp 84

Db 179428 -----CTTTCAAAATCTTTACCTCAATTAAACACGAGAAAGCAATTCATTCAA 179478

Qy 85 GlnLysLysCysAspGluHisLysAlaSerSerProPheSerValAlaLysPheArg 104

179479 GAATCTAGTCAGAACAAATCAAAATGACAGAGAGCG-----179517

Qy 105 ArgTTPAspCysSerLysCysLeuAspLysLeuLysThrSerAspAsnGlyThrAlaPro 124

Db 179518 ---TGACACTATCACTCTCTCTCTCAACAGAGTTCAGACAAACAATACATTT---179571

Qy 125 ArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIleThrPheValArg 144

Db 179572 -----TCGAGTCAGAGTGTAAATGAAGACAAATCTCTCTAGAGATCAAGG 179619

Qy 145 SerThrPheValProAlaSerValGlySerGlnLysValSerProSerThr-----161

Db 179620 ACCTGTGATTTCTCAATCAATTT---ACATGACAAATAGTCCATTACAGATTGAGT 179676

Qy 162 -----GlnSerSerGlnGlyLysAsnAlaAspArgSer---ThrLeuProLysSer 177

Db 179677 GCAGAGATCAACAAATCAAGAGGAAACATCAGAAATCTTCAAAATCTTTACCTCAATTA 179736

Qy 178 ValGlnGluGlyAsn-----AspSerLysCysAsnAlaProSerGlyLysAsn 193

Db 179737 ACAACAGAGAAAGCATTTTCAGATCTAGTCCAGAACAAATCAATCAATGACAGA 179796

Qy 194 GlyAlaAlaGluAlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsn 213

Db 179797 AGAGCGTGGACACTATCACTCTCTC-CTCAACAAAGTTCAGAGACGAACTATCAT 179855

Qy 214 TyrAspValAlaAlaAsnValSerGluAspAsnThrSerVal---AspValGlyAlaLeu 232

Db 179856 TTTTCGAGTCAGAGTGTAAATAT---GAAGACAAATCTCTCTAGAGATCAAGGACATG 179912

Qy 233 ProGluValProGlnIleThrTrpHis-----IleGluValAsnGlyAlaAsp 248

Db 179913 TCGATTTCTCAATCAATTACATGACAAATAGTCCACTTACAGATTTCAGTGCAGAGAT 179972

Qy 249 Gln-----ProProSerThrProLysLeuSerGluValValLeuLysArgAsnGluAsp 266

Db 179973 CAACAAATCAAGAGAAACATCAGAACTTTCAAAATCTTTACCTCAATTAACACAGAA 180032

Qy 267 GluAsnGlyLysThrGluGluThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro 286

Db 180033 GAACATTTCAATTTCAAGATCT---AGTCGAGAACAAATCAATCAATCAAGAGAGCG 180089

Qy 287 AnProMetSerGlyLysGluArgAspGlnValAlaGluGlnCysAsnLeuThr-----304

Db 180090 TGGACACTATCA---ACATCTCTTCTCAACAAAGTTCAGAGACAAACATTCATTTTCG 180146

Qy 305 -----LysAspProLysProValSerProValSerGlyGln 313

Db 180147 AGTCAGATGTATATGAAGACAAATCTCTCTAGAGATCAAGGACACTCTCGATTCT 180206

Qy 314 LysCysGluGlnIleCysAsnGluProCysGluGluValValLeuLysArgSerLys 333

Db 180207 CAATCAATTTACATCAATGACAAATAGTCCACTTACAGATTTCAG-TGCAGAGATCAACAAA 180265

Qy 334 SerLysArg-----LysThrAspLysLysLeuMetLysLysGlnGlnHis 348

Db 180266 TCAGAGAGAAACATCAGAACTTTCAAAATCTTACCTCAATTAACACAGAGAAAGAT 180325

Qy 349 -----SerLysLysArgThrAlaGlnAlaAspVal 358

Db 180326 TTCATTTCAAGATCTAGTGCAGACAAATCAAAATGACAGAGAGCGCTGCAGACTATC 180385

Qy 359 SerAspAlaLysLeuCysArgArgLysProLysLysValArgLeuLeuSerGluIle 378

Db 180386 AACATCTCTTCTCAACAAAGTTCAGAGCAAAAC---ATTACATTTTCGAGTCAGAGTGA 180443

Qy 379 AsnAlaAsnGln-----ValGluAspSerArg-----SerAspGluValHis 392

Db 180444 ATGAGACAAACTTCTCTAGAGGATACAGGACACTGTCTGATTTCTCAATCAATTA 180503

Qy 393 ArgGluAsnAla-----AlaAspProCysGluAspAspArgSerThrIleProValPro 410

180504 ATGACAAATAGTCCATTACAGATTGAGTCGAGAGATCAACAAATCAAGAGAAACA 180563  
Qy 411 MetGluValSerMetAspIlePro----- 418  
180564 TCAGAACTTTCAAAATCTTTACCTCAATTACACAGAGAAAGCAATTTCAATTTCAAGAA 180623  
Db 419 -----ValSerAsnHisThrValGlyGluAsp-----GlyLeuLys 430  
Qy 180624 TCTAGTCAGAGACAAATCAAAATGACAGAGAGCGGTGGACATATCAACATCTCTCTCT 180683  
Db 431 SerSerLysAsnLysThrLysArgLysTyrSer-----AspValValAspAspGlySer 448  
Qy 180684 CAACACAGTTCAAGACAAACATACATTTTCGAGTCAGAGTGTAATGACAGACAAACT 180743  
Db 449 SerLeuMetAsnTrpLeuAsnGlyLysLysLysLysArgThrGlySerValHisHisThrVal 468  
Qy 180744 TCTCTA-----GAGGATCAAGGACACTGTGCGATTTCTCAATCAATA 180785  
Db 469 AlaHisProAlaGlyAsnLeuSerAsnLysLysValThrProThrAlaSerThrGlnHis 488  
Qy 180786 -----ACATGACAAATAGTCACATTACAGATTTGAGTGCGAGAGATCAA 180830  
Db 489 AspAspGluAsnAsp----- 180890  
Qy 180831 ACAAAATCAAGAGAAACATCAGAACTTTCAAAATCTTTACCTCAATTAAACAACGGAAGAA 180890  
Db 497 GlyLeu-----AspThrAsnMetHisLysThrAspValCysGlnHisValSerGlu 513  
Qy 180891 AGCATTTTCATTTCAAGAACTAGTCAGAACAAATCAAAATCAATGACAGAGCGGTGGACA 180950  
Db 514 IleSerThrGlnArgCysSerSerLysGlyLysThrAlaGlyLeuSerLysGlyLysThr 533  
Qy 180951 CTATCAACATCTCTTTCAACAGTTCAAGACAAACATATATTTTCGAGTCAGAGT 181010  
Db 534 HisSerAlaAlaSerThrLysLysGlyGluSerThrArgAsn-----GlyGlnAsnIle 552  
Qy 181011 GTAAATGAAGACAAAACTCTCTAGGAAATCAAGGACACTGTGCGATTTCTCAATCAATA 181070  
Db 553 HisVal-----LeuSerAlaGluAspGlnCysGlnMetGluThr 565  
Qy 181071 ACAATGACAAATAGTCCTACTACAGATTGAGTCGAGAGATCAACAAATCAAGAG--- 181127  
Db 566 GluAsnSerValLeuSerHisSerAlaLysValSerProAlaGluHisAspIleGlnIle 585  
Qy 181128 GAAGCATCAGAACTTTCAAAATCTTTACCTCAATTAACACAGAGAAAGCAATTTCTATT 181187  
Db 586 MetSerAspLeuHisGluGln-----LysAsnGlnHisGluArg----- 593  
Qy 181188 CAAGAAATCTAGTCAGAGAAACAAATCAAAATGACAGAGAGCGGTGGACATATCAACATCT 181247  
Db 594 LeuProLysLysLysLysLysGlnLysLeu-----GluValThrArgGlu 608  
Qy 181248 CTCTCTCAACAGTTCAAGACAAACATATATTTTCGAGTCAGAGTGTAATGAAAGAC 181307  
Db 609 LysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAla----- 625  
Qy 181308 AAAACTTCTAGGAGATACAGGACACTGTGCGATTTCTCAATCAATAACAAAT 181367  
Db 626 -----LysAsnGlnHisGluArg----- 631  
Qy 181368 AGTCTCTTCAGATTGAGTCGAGAGATCAACAAATCAAGAGAAACATCAGAACTT 181427  
Db 632 -----GlnLeuMetThrGluThrAsp----- 638  
Qy 181428 TCAAAATCTTTACCTCAATTAACACAGAGAAAGCAATTTCTATTTCAAGAACTCTAGTGCA 181487  
Db 639 -----Cys 639  
Qy 181488 GAACAAATCAAAATGACAGAGAGCGGTGGACATATCAACATCTTTCTTCAACAAAGT 181547  
Db 640 SerAspIleAsnArgIleGlnSerLysThrThrAlaAspAspCysValIleValala 659  
Qy 181548 TCAAAGACAAACATATATTTTCGAGTCAGAGTGTAATGAGACAAAACTTTCTCTAGAG 181607  
Db 660 AlaLysAspGlySerAspTyrAlaLysSerValPheAspThrAsnSerGlnLysSer 679  
Qy 181608 GATACAGGACACTGTGCGATTTCTCAATCAATCAATGACAAATAGTCCACTTACAGAT 181667  
Db 680 Leu-----AlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAla----- 694  
Qy 181668 TTGAGTCGAGAGATCAAAACAAATCAAGAGAAACATCAGAACTTTTCAAAATCTTTACTT 181727  
Db 695 ---LeuThrThrGlnGluSerProHisProGlnAsnPheGlnSerThrGlnGlnGln 713  
Qy 181728 CAATTACACAGAGAAAGCAATTTCAATTCAGAACTTAGTCGAGACAAATCAATG 181787  
Db 714 ThrHisLeuArgMetGluGluMetValThrIleAlaLysSerProLeuPheSerHis 733  
Qy 181788 ACA-----GAAGAGCGGTGGACACTATCAACATCTCTCTCAATCAAGTTC- 181834  
Db 734 HisAspAspGlnTyrIleAlaGluAlaPro-----ThrGluHisTrpGlyArgLys 750  
Qy 181835 AAAGACAAACATATATTTTCGAGTCAGAGTAAATGAGACAAAACTTCTTAGGG-- 181892  
Db 751 AspAlaLysLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSerProAlaAla 770  
Qy 181893 GATACAGGACACTGTGCGATTTCTCAATCAAGAGAAATAACAAATAGT----- 181940  
Db 771 ThrCysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisVal 790  
Qy 181941 -----CTACTTACAGATTGAGTCGAGAGATCAACAAATCAAGAG--- 181982  
Db 791 MetGlySerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAsp----- 808  
Qy 181983 -----GAAACATCAGAACTTTCAAAATCTTTACCTCAATTAACACAGAGAAAGCAAT 182036  
Db 809 ArgTyrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAla 828  
Qy 182037 TCATTTCAAGAAATCTAGTCGAGAAACAAATCAAAATGACAGAGCGGTGGACATATCA 182096  
Db 829 ThrMet-----GluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValIleu 846  
Qy 182097 ACATCTCTCTCAAAACAGTTCAAGAGACAAACATTAATTTTCGAGTCAGAGTGATAT 182156  
Db 847 TyrProLysGluSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeu 866  
Qy 182157 GAAGACAAAACTTTCTTAGAGGATACAGGACACTGTGCGATTTCTCAATCAATCAATG 182216  
Db 867 AlaSerPheProAsnTyrGlyThrSerArgAsnGlnMetGluSerGlnLeuHisAsn 886  
Qy 182217 ACAATAGTCCACTTACAGATTGAGTCGAGAGATCAACAA----- 182258  
Db 887 SerGlnTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeu 906  
Qy 182259 -----AATCAAGAGAAACATCAGAACTTTCAAAATCTTTTACCTCAATTA 182303  
Db 907 AsnGlyLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHis 926  
Qy 182304 ACAACAGAGAAAGCAATTTCTTTCAAGAACTTAGTCGAGACAAATCAATCAATGACAGAA 182363  
Db 927 ArgProLeuArgProHisProArgValGlyValLeuGlySerLeuLeuGlnLysGluile 946  
Qy 182364 AACCGGTGGACACCATCAACATCTCTCTCAAAACAGTTCAAAAGACAAACATTAATTT 182423  
Db 947 AlaAsnTrpSer-----GluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSer 964  
Qy 182424 TCGAGTCAGAGTGATATGAAGACAA-AACTTCTTAGAGGATACAGGACACTGTGCGAT 182482  
Db 965 ThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGly 984  
Qy 182483 TTCTCATCAAGAGAAATAACAAATAGTCTACTTACAGATTTCAGTCGAGAGATCAAC 182542  
Db 985 MetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAspPhe 1004  
Qy 182543 AAA-----TCAAGAGAGAAACATCAGAACTTTTCAAAATCTTTT 182578  
Db

```
Qy 1005 LeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHis 1024
Db 182579 ACCTCAATTAAACAACAGC----- 182596
Qy 1025 ProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAsn-----Pro 1041
Db 182597 ---AGAAAGCATTTTCATTTTCAGAAATCTAGTCAGAAACAATAATCAATGACAGAGAGCC 182653
Qy 1042 AlaAspPheThrThrIleSerAsnAspAsnGluTyrMetAsp 1055
Db 182654 GTGGACACTATCAACATCTCTTCCTCAACAAGTTCAAGAC 182695
```

Search completed: May 17, 2004, 08:43:32  
Job time : 11717 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 17, 2004, 01:43:21 ; Search time 955 Seconds  
(without alignments)  
4701.932 Million cell updates/sec

Title: US-09-721-114-2

Perfect score: 5526

Sequence: 1 MEIVAVDQEGARVVGTCNML.....NKVPADFTTISNDNRYMDYR 1057

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Dgapop 6.0 , Delxext 7.0

Searched: 3373653 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/Cgn2\_1/USPFCSpool/US09721114/runat 14052004 092250 4074/app query.fasta\_1.1223  
-DB=N\_Geneseq\_29Jan04 -QEXT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOFCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-NO\_XMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXI=7

Database :

N\_Geneseq\_29Jan04: \*  
1: geneseq1980s: \*  
2: geneseq1990s: \*  
3: geneseq2000s: \*  
4: Geneseq2001as: \*  
5: Geneseq2001bs: \*  
6: Geneseq2002as: \*  
7: Geneseq2003as: \*  
8: Geneseq2003bs: \*  
9: Geneseq2003cs: \*  
10: Geneseq2004as: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5526	100.0	3896	7	ABQ77908 Rice OSEM
2	5526	100.0	4310	6	ABA93084 Oryza sat
3	5191	93.9	9455	6	ABA93085 Oryza sat
4	393	7.1	3291	7	ABZ42128 Arabidops
5	298	5.4	8646	3	AAL4873 Nucleotid
6	298	5.4	8648	6	AAL43177 Arabidops
7	298	5.4	8648	7	ABQ77909 Arabidops
8	298	5.4	17341	3	AAL4872 Genomic D

9	298	5.4	17341	6	AAL43176 Arabidops
10	223.5	4.0	5286	5	AAS73156 DNA encod
11	223	4.0	10452	6	ABL68623 Kidney ca
12	223	4.0	10452	6	ABK84422 Human CDN
13	223	4.0	11167	9	ADBS3841 Human pro
14	222.5	4.0	279	6	ABL73385 Corn tass
15	221	4.0	5954	5	AAS80591 DNA encod
16	221	4.0	6143	5	AAS83843 DNA encod
17	220	4.0	3953	4	ABA66556 Human foe
18	220	4.0	3953	4	ABA48746 Human bre
19	220	4.0	3953	4	ABA33721 Probe #12
20	220	4.0	3953	4	AAK40813 Human bon
21	220	4.0	3953	4	AAK15085 Human bra
22	220	4.0	3953	4	ABS40392 Human liv
23	220	4.0	3953	5	AAI07277 Probe #72
24	220	4.0	3953	6	ABS14767 Human gen
25	219	4.0	6201	7	ABX71059 Novel hum
26	217	3.9	3998	4	AAB29827 S cerevis
27	214	3.9	5072	2	AAT71315 cDNA enco
28	214	3.9	5073	2	AAT71314 cDNA enco
29	214	3.9	9642	2	AAT71324 Plasmid p
30	214	3.9	10443	2	AAT71321 Plasmid p
31	214	3.9	13414	2	AAT71321 Plasmid p
32	212.5	3.8	6816	7	ACA19870 Prokaryot
33	211.5	3.8	6783	7	ABT14863 Pathogen
34	211	3.8	7474	2	AAT71320 Plasmid p
35	209	3.8	6813	7	ACF74294 Staphyloc
36	209	3.8	6852	7	ABZ22900 Staphyloc
37	209	3.8	9807	4	ABL06949 Drosophil
38	207	3.7	5495	7	ACA46979 Prokaryot
39	207	3.7	5580	7	ABT14982 Pathogen
40	206.5	3.7	9606	4	ABL12165 Drosophil
41	206.5	3.7	12039	4	ABL12164 Drosophil
42	204.5	3.7	10219	4	ABL20122 Drosophil
43	204	3.7	5355	4	ABL29287 Drosophil
44	204	3.7	8307	4	ABL23161 Drosophil
45	204	3.7	8431	9	ADD29688 Human tum

#### ALIGNMENTS

RESULT 1

ABQ77908  
ID ABQ77908 standard; cDNA; 3896 BP.

XX AC ABQ77908;

XX DT 24-JAN-2003 (first entry)

XX DE Rice OSEM1-encoding cDNA, SEQ ID NO:1.

XX KW Rice; OSEM1; EMF; embryonic flower; plant; floral repressor;  
XX KW reproductive development; flower development; transgenic plant;  
XX KW antisense suppression; transgenic; reciprocal negative interaction;  
XX KW flower meristem identity gene; flowering time; shoot development;  
XX KW seed yield; agriculture; gene; ss.

XX OS Oryza sativa.

XX FH Key Location/Qualifiers

XX CDS 241..3414

XX FT /\*tag= a

XX FT /product= "OSEM1 protein"

XX FN WO200280659-A1.

XX PD 17-OCT-2002.

XX PF 05-APR-2002; 2002WO-US012675.

XX PR 06-APR-2001; 2001US-00828068.

XX (REGC ) UNIV CALIFORNIA.

XX PA





Db 1691 ACACCCACTGGAGTACTCAGCATGATGATGAGTAATGATCTGAAATATGGTCTTGACACA 1740  
Qy 501 AsnMetHisLysThrAspValCysGlnHisValSerGluLysSerThrGlnArgCysSer 520  
Db 1741 AATATGCATAGACAGATGCTGTGCAGCATGTATCAGAAATCTCCACACAGAGGTGCTCA 1800  
Qy 521 SerLysGlyLysThrAlaGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLys 540  
Db 1801 TCAAGGGGAAAAAGCGGGTTTCAGTAAGGGGAAAAACACATTCAGCTGCTAGTACCACA 1860  
Qy 541 TyrGlyGlyGluSerThrArgAsnGlyGlnAsnLysHisValLeuSerAlaGluAspGln 560  
Db 1861 TATGGTGGTGAAGACACAGAAATGTCAGAAATCATACATGCTACTCAGCGCAGAGATCAA 1920  
Qy 561 CysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLysValSerProAlaGlu 580  
Db 1921 TGCAGATGGAAACCGAAATCTGTCTGAGTCACTCGGCAAAAGGTTTCTCCAGCTGAG 1980  
Qy 581 HisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLys 600  
Db 1981 CATGATATCCAAATATGCTGACCTTCATGACGAGAGTCTACCCAAAGAAAGAAAGAG 2040  
Qy 601 GlnLysLeuGluValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIle 620  
Db 2041 CAAAAACCTTGAAGTGACTCGTGAAAAACAGACCATGATGATGACATCCCCCATGGATAT 2100  
Qy 621 ValGluLeuLeuAlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSer 640  
Db 2101 GTGAGCTGCTAGCTAAACACCATGACGAGGAGCTTATGACTGAGACTGATGTCT 2160  
Qy 641 AspIleAsnArgIleGlnSerLysThrThrAlaAspAspCysValIleValAlaAla 660  
Db 2161 GACATCAACCGTATTCAATCCAGACAACTGCTGATGATGATTTGTGTAATAGTAGCTGCC 2220  
Qy 661 LysAspGlySerAspThrAlaSerSerValPheAspThrAsnSerGlnGlnLysSerLeu 680  
Db 2221 AAGGATGGTTCAGATTATGATCATCAAGTGTGTTGACACATTAATCCCAACAGAGTCTCTG 2280  
Qy 681 AlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGluSer 700  
Db 2281 GCATCCCAAGTACACAGAGAGGTACAGGCTCATTTGGCATTTGACCAACACAGAGTCT 2340  
Qy 701 ProHisProGlnAsnPheGlnSerThrGlnGlnGlnThrHisLeuArgMetGluGlu 720  
Db 2341 CCACATCCTCAGAACTTTTCACTTACTTACTCAGGAAACAGCAGACACATTTGCGGAGTGAAGAA 2400  
Qy 721 MetValThrIleAlaAlaSerSerProLeuPheSerHisHisAspAspGlnTyrIleAla 740  
Db 2401 ATGCTCACTATTGCTGAGCTCACCCTATTTTTCATATCATGATGATGATATTTGCT 2460  
Qy 741 GluAlaProThrGluHisLysGlyArgLysAspAlaLysLysLeuThrTrpGluGlnPhe 760  
Db 2461 GAAGCACCACCTGAACATTTGGGGCGGTAAAGCAGCAGCAAGAGCTAAAGTGGAGCAATTT 2520  
Qy 761 LysAlaThrThrArgAsnSerProAlaAlaThrCysGlyAlaGlnPheArgProGlyIle 780  
Db 2521 AAGGCCCACTACAGAAATTTCCAGCAGCAGCAATGCTGCTCAATTTAGACCTGGTATC 2580  
Qy 781 GlnAlaValAspLeuThrSerThrHisValMetGlySerSerSerAsnTyrAlaSerArg 800  
Db 2581 CAACGAGTTGACTTCTTACTCATGTCATGAGTCTTCCAGCAATTAATGATCTCGC 2640  
Qy 801 GlnProValIleAlaProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAla 820  
Db 2641 CAACCACTAATTTGGCCCACTGGACCGCTATGCTGAAAGAGCGGTAAACAGGTTCCATGCA 2700  
Qy 821 ArgAsnPheProSerThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsn 840  
Db 2701 AGAAATTTCCAGACCAATAGCAACCATGAGCAGAGTAAAGTTATGTCGAGGAAT 2760  
Qy 841 AlaGlyGlnValValLeuTyrProLysGluSerMetProAlaThrHisLeuLeuArgMet 860  
Db 2761 GCTGGACAAGTAGTCTTGTATCTCTAAAGAAATCCATGCTCGCAGCATCTTCTCAGAAATG 2820

Qy 861 MetAspProSerThrLeuAlaSerPheProAsnTyrGlyThrSerSerArgAsnGlnMet 880  
Db 2821 ATGGATCCATCAACATTAGCAAGCTTCCCACACTATGGAACCTTAGCAGGAACCCAGATG 2880  
Qy 881 GluSerGlnLeuHisAsnSerGlnTyrAlaHisAsnGlnTyrLysGlySerThrSerThr 900  
Db 2881 GAGTCTCAACTTCATATCTTCAGTATGACATATACATACATACATACATACATACAT 2940  
Qy 901 SerTyrGlySerAsnLeuAsnGlyLysIleProLeuThrPheGluAspLeuSerArgHis 920  
Db 2941 TCATATGGCAGTAACCTGAATGGAAGATTCCTATTCGACATTCGAGACTTATCACGCGCAT 3000  
Qy 921 GlnLeuHisAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlySer 940  
Db 3001 CAGCTGCATGATCTGCACAGACCTTTACGGCCCATCTCTAGAGTTGGTGTGCTTGGCTCC 3060  
Qy 941 LeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLys 960  
Db 3061 TTCTCTCAGAGGAATATGCAAACTGCTCGGACACTGTGGCACACAACTCTGTTATAAG 3120  
Qy 961 LeuGlyValSerThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAla 980  
Db 3121 TTAGGAGTGTCAACAGGAATAACATCGCATGATGAAACAGAAAGGAACATTTTGAAGCC 3180  
Qy 981 LeuAsnSerGlyMetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSer 1000  
Db 3181 CTGAATTCCTGGAATGTTTTCAGCAAAATGGAATGCAATTCAGTTGGTCTGTTAGCTCC 3240  
Qy 1001 SerAlaAspPheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGly 1020  
Db 3241 AGTGCAGATTTTATATCAGCGAGAAACAGCATAGCTCAATCTTTGGACCAGAGCAAGGGT 3300  
Qy 1021 LysMetValHisProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAsn 1040  
Db 3301 AAAATGTTTATCCCTTGGATCGTTTGTGAGACAGATATCTGTTATTAATACTACAGAAC 3360  
Qy 1041 ProAlaAspPheThrThrIleSerAsnAspAsnGluTyrMetAspTyrArg 1057  
Db 3361 CCAGCTGATTTTACTACAATCAGTAACGATAACGAGTATATGATTAACCGC 3411

RESULT 2  
ABA93084  
ID ABA93084 standard; cdna; 4310 BP.  
XX  
AC ABA93084;  
XX  
DT 11-APR-2002 (first entry)  
XX  
DZ Oryza sativa brassinosteroid response related protein encoding cDNA.  
XX  
KW Oryza sativa; rice; plant; brassinosteroid response; control;  
XX signal transfer system; brassinosteroid hormone; growth promotion;  
XX increased yield; quality improvement; ripeness promotion;  
XX stress relaxation; chemical resistance; gene; ss.  
XX  
OS Oryza sativa.  
XX  
FH Key Location/Qualifiers  
FT CDS 655..3828  
FT /tag= a  
FT /product= "brassinosteroid response related protein"  
XX  
FN JP2001327287-A.  
XX  
PD 27-NOV-2001.  
XX  
PF 19-MAY-2000; 2000JP-00149106.  
XX  
PR 19-MAY-2000; 2000JP-00149106.  
XX  
XX (NORQ) NORINSUISANSHO NOGOYO SEIBUTSU SHIGEN.  
PA (SEIB-) SEIBUTSUKEI TOKUTEI SANGYO GIJUTSU.  
PA

XX WFI; 2002-135739/18.  
DR P-PSDB; ABB05428.  
XX  
PT New gene involved in brassinosteroid responses useful for controlling the  
PT effects such as growth promotion, increased yield, quality improvement,  
PT ripeness promotion, stress relaxation and chemical resistance.  
XX  
XX  
PS Disclosure; Page 6-10; 19pp; Japanese.  
XX  
CC The present invention describes a polynucleotide encoding a plant gene  
CC which can control the signal transfer system of brassinosteroid hormone.  
CC The polynucleotide can be used for controlling the effects such as growth  
CC promotion, increased yield, quality improvement, ripeness promotion,  
CC stress relaxation and chemical resistance. The present sequence encodes a  
CC brassinosteroid response related protein isolated from rice (*Oryza*  
CC *sativa*), from the present invention  
XX  
SQ Sequence 4310 BP; 1237 A; 986 C; 1022 G; 1065 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 0 Length: 4310  
Score: 5526.00 Matches: 1057  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
  
US-09-721-114-2 (1-1057) x ABA93084 (1-4310)  
QY 1 MetGluLeuValAlaValAspGlnGluGlyAlaArgValValGlyThrAsnCysMetLeu 20  
DB 655 ATGGAGATTGTTCAGTAGATCAGGAGGAGCTGCTGTTTGGGAGCAACTGTATGCTT 714  
QY 21 AlaArgGlyGlyThrGlyAlaValAlaProValLeuGluLeuThrAlaThrProArgGln 40  
DB 715 GCTGCTGGTGAAGTCTGTGTAGCCAGTGTGGAGCTGACAGGACGCTCGTCAG 774  
QY 41 AspAlaAlaGluAlaGlyValAspGluProAlaGlnHisGlnCysGluHisPheSer 60  
DB 775 GATCAGCCGCTGAAGCTGGTGTAGACCAACCGGACCAACCACTCCGAGCAATTCCTCC 834  
QY 61 IleArgGlyTyrValAlaLeuLeuGlnLysLysAspProLysPheCysSerLeuSerArg 80  
DB 835 ATAGAGGGTATGTGTCTCTTCAGAGAGGATCCAAATTCCTCTCTATCTCGG 894  
QY 81 IlePheHisAspGlnLysLysCysAspGluHisLysAlaSerSerProPheSerVal 100  
DB 895 ATTTTCCATGACCAAGAAATGTGATCAACACAAAGCTAGTTCAGGCCCATTTCTGTA 954  
QY 101 AlalysPheArgGlyTyrAspCysSerLysCysLeuAspLysLeuLysThrSerAspAsn 120  
DB 955 GCAAGTTCACCATGGATGCTCGAAGTGTCTGGATAGTTGAAATCTTCAGATAT 1014  
QY 121 GlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerile 140  
DB 1015 GGAACAGCACCAAGAACTTTCCCGAAAGCAGAAATGSCACAGTGTGCTCCATC 1074  
QY 141 ThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysValSerProSer 160  
DB 1075 ACATTTGTCGGAGCACTTTTGTGCTGTAGTGTGTTCCCAAGAGTGTCTCCTAGC 1134  
QY 161 ThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGlnGlu 180  
DB 1135 ACACAAATCATCTCAAGGGAGAGATGCTGATAGATCAACTCTCTCCAAAGAGTGTGCAAGAA 1194  
QY 181 GlyAsnAspSerLysCysAsnAlaProSerGlyLysAsnGlyAlaAlaGluAlaAsnThr 200  
DB 1195 GGCATGACTCCAAATGCAATGCGCTTCTGGCAGAGATGGAGTGTCTGGAGGCCCAATCT 1254  
QY 201 AspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAlaAlaAsnVal 220  
DB 1255 GATTCCCAATGAAGATTTGCAAGGGCCGACCCCAAAATATGATGTGGCAGCAATGTC 1314

QY 221 SerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIleThrTrp 240  
DB 1315 TCTGAGGACACACTTCTGTTGATGTTGGGGCTTTACCTGAAGTTCCCCAGATTACATGG 1374  
QY 241 HisIleGluValAsnGlyAlaAspGlnProProSerThrProLysLeuSerGluValVal 260  
DB 1375 CACATAGAAGTAATATGGTGCAGATCACTCCATCCACTCCAAAACCTTTCTGAAGTGTTC 1434  
QY 261 LeuLysArgAsnGluAspGluAsnGlyLysThrGluGluThrLeuValAlaGluGlnCys 280  
DB 1435 CTCAAAAGAAATGAGATGAAATGGAATAAACTGAAGAGACTTTGTTGCTGAGCAGTGC 1494  
QY 281 AsnLeuThrLysAspProAsnProMetSerGlyLysGluArgAspGlnValAlaGluGln 300  
DB 1495 AATTATGACCAAGATCTCTAAACCAATGCTCGAAAGAACGATGATCAGGTTTCTGAGCAG 1554  
QY 301 CysAsnLeuThrLysAspProLysProValSerGlyGlnLysCysGluGlnIleCysAsn 320  
DB 1555 TCAATATTTGACCAAGATCCGAAACCGTGTCTGGGCGAATAATGTGACAGATCTGCAAT 1614  
QY 321 GluProCysGluGluValValLeuLysArgSerSerLysSerLysArgLysThrAspLys 340  
DB 1615 GAGCCATGTGAAGAAAGTTGTTCTCAAAAGAGCTCCAAATCTAAGAGGAAGACGATAG 1674  
QY 341 LysLeuMetLysLysGlnGluHisSerLysLysArgThrAlaGlnAlaAspValSerAsp 360  
DB 1675 AAGTTGATGAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1734  
QY 361 AlaLysLeuCysArgArgLysProLysLysValArgLeuLeuSerGluIleLeuAsnAla 380  
DB 1735 GCATAAGCTTTGTCGAGAAAGCCAAAAGAGTGGCGCTTCTATCAGAAATATAAATGCT 1794  
QY 381 AsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspProCys 400  
DB 1795 AACCAAGTTGAGGATTCAGAAAGTTCATCGTGAATAATGCCCGCTGATCCCTGT 1854  
QY 401 GluAspAspArgSerThrIleProValProMetGluValSerMetAspIleProValSer 420  
DB 1855 GAGGATGATAGAGTACCATCCCGTCCGATGGAGTAAGCATGGATATTCTGTGTAGC 1914  
QY 421 AsnHisThrValGlyGluAspGlyLeuLysSerSerLysAsnLysThrLysArgLysTyr 440  
DB 1915 AACCATACAGTGGGAGAGATGGGTAAATCAAGTAAGAACACAGACAAAACGCAAAATAC 1974  
QY 441 SerAspValValAspAspGlySerSerLeuMetAsnTyrLeuAsnGlyLysLysLysArg 460  
DB 1975 TCTGATGTTGTAGATGATGGATCATCATTAATGAACTGGCTGATGGAATAAGAAAGA 2034  
QY 461 ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
DB 2035 ACTGGAAGTGTGCATCACACAGTGTCTCATCCAGCTGGGAATTTGAGCAACAAAAAGTG 2094  
QY 481 ThrProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAspThr 500  
DB 2095 ACACCCACTGCCAGTACTCAGCATGATGAGATGATGATGATGATGATGATGATGATGAT 2154  
QY 501 AsnMetHisLysThrAspValCysGlnHisValSerGluIleSerThrGlnArgCysSer 520  
DB 2155 AATATCATTAAGACAGATGCTGTGTCAGCATGTATCAGAAATCTCCACACAGAGTGTCA 2214  
QY 521 SerLysGlyLysThrAlaGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLys 540  
DB 2215 TCAAGGGGAAACACGCGGTTTGAAGTAAAGGGGAAACACATTCAGCTGCTAGTACCAA 2274  
QY 541 TyrGlyGlyGluSerThrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGln 560  
DB 2275 TATGTTGTTGAAGACACAGAAATGTGTGAGACATACATGATCTCAGCGCAGAAAGTCAA 2334  
QY 561 CysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLysValSerProAlaGlu 580  
DB 2335 TGCCAGATGAAACCCGAAACCTCTGTCTGAGTCACTCGGCAAGAGGTTCCTCCAGCTGAG 2394

QY 581 HisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLys 600  
 DB 2395 CATGATATCCAAATATATGCTGACCTTCATGAGCAGAGTCTACCCAGAGAGAGAGAG 2454  
 QY 601 GlnLysLeuGluValThrArgGluGlnThrMetIleAspAspIleProMetAspIle 620  
 DB 2455 CAATAAATTGAGTGACTCTGTGAATAACAGACCAATGATGATGATCCCATGATATT 2514  
 QY 621 ValGluLeuLeuAlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSer 640  
 DB 2515 GTTGAACCTGCTAGCTAAACACAGCATGAGAGGAGCTTATGACTGAGAGCTGATTGTCT 2574  
 QY 641 AspIleAsnArgIleGlnSerLysThrAlaAspAspAspCysValIleValAlaAla 660  
 DB 2575 GACATCAACCGATATTCAATCCAGACCACTGCTGATGATGATGATGATGATGATGCTG 2634  
 QY 661 LysAspGlySerAspThrAlaSerSerValPheAspThrAsnSerGlnGlnLysSerLeu 680  
 DB 2635 AAGGATGGTTCAGATTATGATCAATCAAGTGTGTTGACACTAATTCCTCCACAGAGCTCTG 2694  
 QY 681 AlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGluSer 700  
 DB 2695 GCATCCCAAGTACACAGAGAGAGTTCAGGGTCAATTTGGCATTTGACCCACAGAGTCT 2754  
 QY 701 ProHisProGlnAsnPheGlnSerThrGlnGluGlnGlnThrHisLeuArgMetGluGlu 720  
 DB 2755 CCATCTCTCAACTTTCAGTCTACTCAGGACAGCAGACACATTTGGCGATGAGAGAA 2814  
 QY 721 MetValThrIleAlaLeuSerSerProLeuPheSerHisHisAspAspGlnThrIleAla 740  
 DB 2815 ATGGTCACTATTGCTGAAGCTCACCCTATTTTTCATCATCATGATCATCATGATATTGCT 2874  
 QY 741 GluAlaProThrGluHisTrpGlyArgLysAspAlaLysLysLeuThrTrpGluGlnPhe 760  
 DB 2875 GARGCACCACCTGAACTGAGCTGGGCGGTGAAGACCAAGAGCTAACTGGGAGCAATTT 2934  
 QY 761 LysAlaThrThrArgAsnSerProAlaAlaThrCysGlyAlaGlnPheArgProGlyIle 780  
 DB 2935 AAGGCCACTACAAGAAATTTCTCCAGCAGCAACATGTGTGCTCAATTTAGACCTGGTATC 2994  
 QY 781 GlnAlaValAspLeuThrSerThrHisValMetGlySerSerSerSerSerSerSerArg 800  
 DB 2995 CAAGCAGTTGACTTGCATCTACTCATGTCATGCGATCTTCCAGCAATTTATGCACTCTGC 3054  
 QY 801 GlnProValIleAlaProLeuAspArgTyraAlaGluArgAlaValAsnGlnValHisAla 820  
 DB 3055 CAACCAAGTAATGGCCACTGACCGCTATGCTGAAGAGCGGTTACCAAGTCCATGCA 3114  
 QY 821 ArgAsnPheProSerThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsn 840  
 DB 3115 AGAAATTTTCCACAGCACAAATAGCAACCATGGAGCGAGTAGTATTATGATCGGAGAAAT 3174  
 QY 841 AlaGlyGlnValValLeuTyrrProLysGluSerMetProAlaThrHisLeuLeuArgMet 860  
 DB 3175 CTGACCAAGTAGTCTTGATCTTAAGAAATCCATGCTGGAGCGCATCTTCTGAGATG 3234  
 QY 861 MetAspProSerThrLeuAlaSerPheProAsnTyrrGlyThrSerSerArgAsnGlnMet 880  
 DB 3235 ATGGATCCATCAACATTAGCAAGCTTCCCAACTATGGAACCTTAGCAGGACCAAGATG 3294  
 QY 881 GluSerGlnLeuHisAsnSerGlnTyraHisGlnGlnTyrrLysGlySerThrSerThr 900  
 DB 3295 GAGTCTCAACTTCATAATCTCAGTATGCAATATCATGTAACAAGGATCAACAGCACA 3354  
 QY 901 SerTyrrGlySerAsnLeuAsnGlyLysIleProLeuThrPheGluAspLeuSerArgHis 920  
 DB 3355 TCATATGCGAGTAACCTGAATGGAAGATTCATTTGACATTCGAAGACTTATATCAGGCAT 3414  
 QY 921 GlnLeuHisAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlySer 940  
 DB 3415 CAGCTGCATGATCTGCAGACACCTTACGCCACATCTGAGTGGTGGCTGGCTCC 3474  
 QY 941 LeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrrLys 960

DB 3475 TTCTCTCAGAGGAAATTCGAACTGGTCGAGAACTGGGCACACATCTGTTATPAG 3534  
 QY 961 LeuGlyValSerThrGlyIleThrSerHisGlnMetAspArgLysGluHisPheGluAla 980  
 DB 3535 TTAGGAGTGTCAACAGGATAACATCGATCAGATGAACAGAAAGAACATTITGAGCC 3594  
 QY 981 LeuAsnSerGlyMetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSer 1000  
 DB 3595 CTGAATCTTGGAATGTTTTCAGCAAAATGGAATGCAATTCAGTTGGGTTCTGTTAGCTCC 3654  
 QY 1001 SerAlaAspPheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGly 1020  
 DB 3655 AGTGCAGATTTTATCAGCGAGGACAGACATAGCTCAATCTTGGACCAGAGCAAGGT 3714  
 QY 1021 LysMetValHisProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAsn 1040  
 DB 3715 AAAATGGTTTCATCCCTTGGATCGGTTTGTGAGACAGGATATCTGTATACTAACAGAAC 3774  
 QY 1041 ProAlaAspPheThrThrIleSerAsnAspAsnGluTyrrMetAspTyrrArg 1057  
 DB 3775 CCAGCTGATTTTACTACATCATCAGTAACGATACGATATATGATTAACCCG 3825  
 RESULT 3  
 ABA93085  
 ID ABA93085 standard; DNA; 9455 BP.  
 XX ABA93085;  
 AC ABA93085;  
 DT 11-APR-2002 (first entry)  
 XX  
 DE Oryza sativa brassinosteroid response related protein related DNA.  
 XX  
 KW Oryza sativa; rice; plant; brassinosteroid response; control;  
 KW signal transfer system; brassinosteroid hormone; growth promotion;  
 KW increased yield; quality improvement; ripeness promotion;  
 KW stress relaxation; chemical resistance; gene; ds.  
 XX  
 OS Oryza sativa.  
 XX  
 PN JP2001327287-A.  
 XX  
 PD 27-NOV-2001.  
 XX  
 PF 19-MAY-2000; 2000JP-00149106.  
 XX  
 PR 19-MAY-2000; 2000JP-00149106.  
 XX  
 PA (NORQ) NORINSUISANSHO NOGYO SEIBUTSU SHIGEN.  
 PA (SEIB-) SEIBUTSUKI TOKUKEI SANGYO GIJUTSU.  
 XX  
 DR WPI; 2002-135739/18.  
 XX  
 XX New gene involved in brassinosteroid responses useful for controlling the  
 PT effects such as growth promotion, increased yield, quality improvement,  
 PT ripeness promotion, stress relaxation and chemical resistance.  
 XX  
 PS Disclosure; Page 13-16; 19pp; Japanese.  
 XX  
 CC The present invention describes a polynucleotide encoding a plant gene  
 CC which can control the signal transfer system of brassinosteroid hormone.  
 CC The polynucleotide can be used for controlling the effects such as growth  
 CC promotion, increased yield, quality improvement, ripeness promotion,  
 CC stress relaxation and chemical resistance. The present sequence  
 CC represents a DNA sequence related to a brassinosteroid response related  
 CC protein isolated from rice (Oryza sativa), from the present invention  
 XX  
 SQ Sequence 9455 BP; 2571 A; 2001 C; 2090 G; 2793 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 0 Length: 9455  
 Score: 5191.00 Matches: 1055  
 Percent Similarity: 68.06% Conservative: 0

Best Local Similarity:	68.06%	Mismatches:	2
Query Match:	93.94%	Indels:	495
DB:	6	Gaps:	3
US-09-721-114-2 (1-1057) x ABA93085 (1-9455)			
Qy	1	MetGluIleValAlaValaspGlnGluGlyAlaargValValGlyThrAsnCysMetLeu	20
Db	4223	ATCGAGATTCTTGCAGTAGATFCAGGAGGAGCTGCTGTTGTGGACGAACATGTATGCTT	4282
Qy	21	AlaArgGlyGlyThrGlyAlaValAlaProValIleuGluLeuThrAlaThrProArgGln	40
Db	4283	GCTCGTGGTGAACCTGCTGCTAGCCAGCTGTTGGAGCTGACAGCGACCTGCTCAG	4342
Qy	41	AspAlaAlaGluAlaGlyValaspGluProAlaGlnHisGlnCysGluHisPheSer	60
Db	4343	GATCGACCGCTGAAGCTGGTGTAGACGAACCGGCACACCAATGCGAGCATTTCTCC	4402
Qy	61	Ile-----	61
Db	4403	AT-AAGGTAAATCATTTTCTGTATTTCCAAATTCAGTATCGGTTGTGGATGAATATGAA	4461
Qy	61	-----	61
Db	4462	TCGGCATGTCATGCCATATTGCACTGTTTGATGGAGAGATGATGATGATGCTGTTTTTG	4521
Qy	61	-----	61
Db	4522	CACAGTTTGCTTTGGGACTTATATGTCATCTGTTTTGTACGATCGTATACACTGGGTC	4581
Qy	61	-----	61
Db	4582	GACATGCTTATGACTTCTGGTTCGATTTAGGAAGTCAATACATCCACTACTAGCTCTATAT	4641
Qy	61	-----	61
Db	4642	CTAGCCATGTGAACCTCATTTATGCCATAGCAGCTAGCAGGCTAGCAGCAAAAATATA	4701
Qy	61	-----	61
Db	4702	TATAATTTTGCATATATATGTTGGTGTTCATGATCTTTACTCTAGCTACATCCATTA	4761
Qy	61	-----	61
Db	4762	ATATCTTCAATGTATGAATCTGAGCAGATGATTTGAGTGCTACACATATGCAATGCTGT	4821
Qy	61	-----	61
Db	4822	ATGTGTTTCANTAGGTGTTGATCATATTTGTTGTGTTGGGTGCGCATGCAATTATT	4881
Qy	61	-----	61
Db	4882	CAGGCCATGCTGTAGGCTGTAGCTAGATATTTGTGTTTGTATATTTTCTGTGTGAACA	4941
Qy	61	-----	61
Db	4942	GCTGATTACTAATGAATTAACCTTTTGGGTACATCATATATGGGCCCTACATTTT	5001
Qy	61	-----	61
Db	5002	TGTAAATCATTTTCTTTGTGCTCTGAGGTTGAGCATATAAACTTTTTTATCATAGCATGTT	5061
Qy	61	-----	61
Db	5062	TACATCCTTAGGAGATCTTTAGAACTGATGGTTTCTTCATATTTGCAATTAAGTTTGTG	5121
Qy	61	-----	61
Db	5122	ATAGTCCATTATTTTAAAGCCTTTTCAATTTGTTTAGAGATTTCTAGAGATGATATATAT	5181
Qy	61	-----	61
Db	5182	CAACCATAGACTTTGTCACGTTTTGGTTTAACTACTTTCCTAGAACAAATTAGATTAATTT	5241

QY	61	-----	61
DB	5242	TTGTAGTTTATCTGTCACTGCTATTTGTATATCTTTGAATTCAAACTGCAATCTTAGA	5301
QY	61	-----	61
DB	5302	TTATCTTGAAGGTCTCTCTTTCTGTGACCTGTACAGCTATGTATGAAATGCCTACCTCCC	5361
QY	61	-----	61
DB	5362	AGCATCCTTTAGATTATGTAGGGCCTTTTCTGAGTTTATCATGTTGTATATGACTGAAGC	5421
QY	61	-----	61
DB	5422	ACGCAATGTGCTATATATATATGTGCCATCTTTATATGATAATCTTATTTTCTTGT	5481
QY	62	---ArgGlyTyrValAlaLeuLeuGlnLysLeuAspProLysPheCysSerLeuSerArg	80
DB	5482	ACCAGAGGTATGTGTCTCTCTTCAAGAAGGATCCAAATTTCTGCTCTCTATCTCGG	5541
QY	81	IlePheHisepGlnLysLysCysAspGluHisLysAlaSerSerSerProPheSerVal	100
DB	5542	ATTTTCATGACCCGAAATAATGTATGATGACACAAAGCTAGTTCAGCCCATTTCTGTA	5601
QY	101	AlaLysPheArgArgTyrAspCysSerLysCysLeuAspLysLeuLysThrSerAspAsn	120
DB	5602	GCAAAATTTGACGATGGGATTTGCTCGAAGTGTGATTAAGTTGAAATCTTCAGATAAT	5661
QY	121	GlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIle	140
DB	5662	GGACAGACCAACAGACTCTTCCCGCAAGACGAATGGCAAGTGTATGTTTCTCATC	5721
QY	141	ThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysValSerProSer	160
DB	5722	ACATTTGTTGGAGCACTTTTGTGTGCTGTAGTGTGGTTCCCAAAAGTGTCTCTAGC	5781
QY	161	ThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGlnGlu	180
DB	5782	ACACAAATCATCTCAAGGGAAGATGCTGTATAGTCAACTCTTCCAAAGAGTGTCAAGAA	5841
QY	181	GlyAsnAspSerLysCysAsnAlaProSerGlyLysAsnGlyAlaAlaGluAlaAsnThr	200
DB	5842	GGCAATGACTCCAANTGCATGGCCCTTCTGGCAAGAATGGAGCTGCTGAGGCCAATACT	5901
QY	201	AspSerProMetLys-----	205
DB	5902	GATTCACCAATGAA-AGGTATGTGTAGATGTAGAGCCTTTCAAATTCCTAAGTAGGATTTT	5960
QY	206	-----AspLeuGlnGlyProAla	211
DB	5961	ATTTAAGGTATAGATAATAACTATGTTGTGTGATTTTCTCAGATTTGCAAGGGCCAGCC	6020
QY	212	GlnAsnTyrAspValAlaAlaAsnValSerGlnAspAsnThrSerValAspValGlyAla	231
DB	6021	CAAAATTTATGATGTGGCAGCAAAATGTCTCTGAGGACCAACACTTCTGTGTATGTGGGGCT	6080
QY	232	LeuProGluValProGlnIleThrTyrHisIleGluValAsnGlyAlaAspGlnProPro	251
DB	6081	TTACTCTGAAGTTCCCGAGTTTACATGGCAATAGAGTAAATGGTGCAGATCAACCTCCA	6140
QY	252	SerThrProLysLeuSerGluValValLeuLysArgAsnGluAspGluAsnGlyLysThr	271
DB	6141	TCCACTCCAAACTTTCTGAAGTGGTCTCTCAAGAAATGAAGATGAAATGGAATAACT	6200
QY	272	GluGlnThrLeuValAlaGluGlnCysAsnLeuThrLysAspProAsnProMetSerGly	291
DB	6201	GAAGAGACTCTTGTGTGTGACGATGGCAATTTGACCAAGATCTCTAACCCCATGTTCTTGG	6260
QY	292	LysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLysProValSer	311
DB	6261	AGCAACCTGATCAGTTTGTGTGACGATGCAATTTGACCAAGATCCGAACACAGTGTCT	6320

QY 312 GlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluValValLeuLysArgSer 331  
Db 6321 GGGCAGAAATGTGAGCAGATCTCAATGAGCCATGTGAAGAAGTTGTCTCAAAAGAAGC 6380  
QY 332 SerLysSerLysArgLysThrAspLysLysLeuMetLysLysGlnGlnHisSerLysLys 351  
Db 6381 TCCAAATCTAAGAGAGACGGATAGAAAGTTGATGAGAGACGACGACACCAAGAA 6440  
QY 352 ArgThrAlaGlnAlaAspValSerAspAlaLysLeuCysArgArgLysProLysLysVal 371  
Db 6441 CGCACTGCCAGCGTGATGTTTCAGATGCAAAAGCTTGTGCGGAAAGCCAAAAGGTTG 6500  
QY 372 ArgLeuLeuSerGluIleIleAsnAlaAsnGlnValGluAspSerArgSerAspGluVal 391  
Db 6501 CGGCTTCATCAGAAATATTAATATCTAACAGGTTGAGANTTCTAGAAGTGACGAGTT 6560  
QY 392 HisArgGlnAsnAlaAlaAspProCysGluAspAspArgSerThrIleProValProMet 411  
Db 6561 CATCGTGAAATGCCGTGATCCCTGTGAGCATGATAGAAAGTACCACCCCGTCCCGATG 6620  
QY 412 GluValSerMetAspIleProValSerAsnHisThrValGlyGluAspGlyLeuLysSer 431  
Db 6621 GAAGTAAGCATGGATATCTCTGTTAGCAACATACAGTGGGAGAAAGTGGTTAAATCA 6680  
QY 432 SerLysAsnLysThrLysArgLysTyrSerAspValValAspAspGlySerSerLeuMet 451  
Db 6681 AGTAAGAACCAAGCAAAACCAATACTCTGATGTTGTAGATGATGATCATCATATG 6740  
QY 452 AsnThrLeuAsnGlyLysLysArgThrGlySerValHisThrValAlaHisPro 471  
Db 6741 AACTGGCTGAATGCAAAAAAGAAAGAACTGGAGTGTGATCACACAGATTGCTCATCCA 6800  
QY 472 AlaGlyAsnLeuSerAsnLysLysValThrProThrAlaSerThrGlnHisAspAspGlu 491  
Db 6801 GCTGGGANTTGGACACAAAGAGTGCACCCACTCCGAGTACTCAGCATGATGATGAG 6860  
QY 492 AsnAspThrGluAsnGlyLeuAspThrAsnMethHisLysThrAspValCysGlnHisVal 511  
Db 6861 AATGATACTGAAATGGTCTTGGACACAAATATGATTAAGACAGATGCTCTCAGCATGTA 6920  
QY 512 SerGluIleSerThrGlnArgCysSerSerLysGlyLysThrAlaGlyLeuSerLysGly 531  
Db 6921 TCAGAAATCTCCACAGAGTGCTCATCAAGGEGGAACAGCCGGTGTGTAGTAAAGGG 6980  
QY 532 LysThrHisSerAlaAlaSerThrLysTyrGlyGlyGluSerThrArgAsnGlyGlnAsn 551  
Db 6981 AAAACACATTGAGCTGTAGTACCAATATGGTGGTGAAGACACCAAGAAATGGTCAGAAC 7040  
QY 552 IleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrGluAsnSerValLeuSer 571  
Db 7041 ATACATGTACTCAGCGCAGAGATCAATGCCAGATGGAAACCGGAAACTCTGTTCTGAGT 7100  
QY 572 HisSerAlaLys----- 575  
Db 7101 CACTCGGCAAGGTACGAAATTTTGTGATCATGAGGAATTTTGTCTTTTAAATTGACTG 7160  
QY 575 ----- 575  
Db 7161 AATCAACATTTATCTGTATGAAGGAATAATATGGTGATCAACAAATGTTAAGAAATATGC 7220  
QY 575 ----- 575  
Db 7221 ATCAAGTCTTATTATATGCTTTCCACTGTCTCTTCTTACTTATGTTTTCATCTCTT 7280  
QY 575 ----- 575  
Db 7281 TTGTGTGTGCGTGATGTGATGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 7340  
QY 575 ----- 575  
Db 7341 ACGTGGTGGCGCAATATCTTTTATAGCTCATATATATGATGATGATGATGATGATGATGATG 7400  
QY 576 -----ValSerProAlaGluHisAspIleGlnIleMetSerAs 588

Db 7401 TTTTCTCTCATTTCTCATCTCAGGTTTCTCCAGCTGAGCATGATATTCAAATTTATGCTGA 7460  
QY 588 pLeuHisGluGlnSerLeuProLysLysLysLysLysGlnLysLeuGluValThrArgG1 608  
Db 7461 CCTTCATGACGAGTCTACCCCAAGAAAGAAAGCAAAACCTTGAAGTGAAGTGAAGTGA 7520  
QY 608 uLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysSerG1 628  
Db 7521 AAAACAGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7580  
QY 628 nHisGluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgIleGlnSerLys 648  
Db 7581 GCATGAGAGCGAGCTTATGACTGAGACTGATTTGTTCTGACATCAACCGTATTCATCCAA 7640  
QY 648 sThrThrAlaAspAspAspCysValIleValAlaLysAspGlySerAspTyrAlase 668  
Db 7641 GCAACTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7700  
QY 668 rSerValPheAspThrAsnSerGlnGlnLysSerLeuAlaSerGlnSerThrGlnLysG1 688  
Db 7701 AAGTGTGTTTGACACTAATTCACACAGAGTCCCTTGGCATCCCAAAAGTACAGAAAGCA 7760  
QY 688 uLeuGlnGlyHisLeuAlaLeuThrThrGlnGluSerProHisProGlnAsnPheGlnse 708  
Db 7761 GTTACAGGGTCAATTTGGCATTTGACACACAGAGTCTCCACATCTCCAGAACTTTCCAGTC 7820  
QY 708 rThrGlnGlnGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerSe 728  
Db 7821 TACTCAGGAACGACGACACATTTGGGATGGAGAAATGGTCACTATTTGCTGCAAGCTC 7880  
QY 728 rProLeuPheSerHisHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTrpG1 748  
Db 7881 ACCACTATTTTCCATCATCATGATGATCAGTATATTTGCTGAAGCACCACCACTGAACATTTGGGG 7940  
QY 748 yArgLysAspAlaLysLeuLeuThrTrpGlnGlnPheLysAlaThrThrArgAsnSerPr 768  
Db 7941 CGGTAAAGGACGCAAGAGACTAACGTTGGGAGCAATTTAAGGCCACTACAGAAATTTCTCC 8000  
QY 768 oAlaAlaThrCysGlyValaGlnPheArgProGlyIleGlnAlaValaAspLeuThrSerTh 788  
Db 8001 AGCAGCAACATGTGGTGTCTCAATTTAGACTGTGTATCCCAAGCAGTGTGACTTGACTTCTAC 8060  
QY 788 rHisValMetGlySerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAs 808  
Db 8061 TCATGTGATGGGATCTTCAGCAATTTATGATCTGCCAACCCAGTAATTTGGCCACTGGA 8120  
QY 808 pArgTyrAlaGluArgAlaValaAsnGlnValHisAlaArgAsnPheProSerThrIleAl 828  
Db 8121 CCCTATGCTGAAGAGCGGTTAACCCAGGTCCATGCAAGAAATTTTCCAAGCACAATAGC 8180  
QY 828 aThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValValLeuTyrPr 848  
Db 8181 AACCATGGAAGCGGTAAGTTATGTGATCGGAAATGTGGACAAGTAGTCTTGTATATCC 8240  
QY 848 oLysGluSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSe 868  
Db 8241 TAAAGAAATCCATGCTTGGAGCGCATCTTCTGAGATGATGATGATGATGATGATGATG 8300  
QY 868 rPheProAsnTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerG1 888  
Db 8301 CTTCCCAACTATGGAACITCTAGCAGAACCAAGATGGAGTCTCAACTTCATTAATTTCTCA 8360  
QY 888 nTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerThrGlySerAsnLeuAsnG1 908  
Db 8361 GTATGACATTAATCAGTACAAAGANTCAACACACATCATATGGCAGTAACTGATGAG 8420  
QY 908 yLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgPr 928  
Db 8421 AAAGATTCCATTGACATTCGAAGACTTATCAGGCGATCATCGCATGATGATGATGATGATG 8480  
QY 928 oLeuArgProHisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIleAlaAs 948





592 GlnSerLeuProIysLysLysLysGlnLysLeuGlnValThrArgGluLysGlnThr 611  
Qy ||||| : : : : :  
Db CATGGAGCACCAAGAGCAAGACTCAGATAGAAAGGAGGACTACGACTGAAGGCAAAAC 1818  
Qy 612 MetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysAsnGlnHisGluArg 631  
Db ||||| : : : : :  
1819 AAC--GATGATATTCCAATGGAGATAGTGAGACTCATGCCCAAAAAACCAAGTACGAGAGG 1875  
Qy 632 GlnLeuMet---ThrGlnThrAspCysSerAspIleAsnArgIleGlnSerLysThr 650  
Qy ||||| : : : : :  
Db 1876 TGTCTTCTCCGACAAAGAGAAAGATGTTAGCAACAAACAGCCATCAACAAGAAACAGCACAC 1935  
Qy 651 AlaAspAspAspCysValIleValAlaAlaLysAspGlySerAspTyrAlaSerSerVal 670  
Qy ||||| : : : : :  
Db 1936 AAATCCAGAAGATGCTCTACTGATTGATCTCAATGAAACCTACGATAACGGGATCTCACTT 1995  
Qy 671 PheAspThrAsnSerGlnGln-----LysSerLeuAlaSerGlnSerThrGlnLysGlu 688  
Db ||||| : : : : :  
1996 GAGGACAAACACATCAAGACCAACAAACCGGTAGTAGCAACGCAAGGAGGGAGAA 2055  
Qy 689 LeuGlnGlyHisLeuAlaLeuThrThrGlnGlnSerProHis----- 702  
Db ||||| : : : : :  
2056 -----CATTTTCTTANGGAAGACAGACACTCTCATGACTTCTTCCCAATAAGT 2106  
Qy 703 -----ProGlnAsn-----PheGlnSerThrGlnGlnGlnGlnThrHis 715  
Db ||||| : : : : :  
2107 CAGCCTTANGTGCCTTCTCGTTTGGGATCTTCTTCCTACCCCAAGAAACCCGA----- 2160  
Qy 716 LeuArgMetGluGluMetValThrIleAlaAspSerProLeuPheSerHisHisAsp 735  
Db ||||| : : : : :  
2161 -----GCCAGCTCCATCCGGTCTTCTTGCTGCACAAAC 2190  
Qy 736 AspGlnTyrIleAlaGluAlaProThrGluHisTyrGlyArgLysAspAlaLysLysLeu 755  
Db ||||| : : : : :  
2191 TGTCACTGGCTCGGGAATTTGCCAACTGTG-----GGTAATCAGAAACCTTCTCCATCC 2244  
Qy 756 ThrTyrGlnGlnPheLysAlaThrThrArgAsnSerProAlaAlaThrCysGlyAla--- 774  
Db ||||| : : : : :  
2245 TCATTTTCGGGTATTACGTGCT-----TGTGATACTTGCACAGAGTGT 2286  
Qy 775 -----GlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisValMetGly 792  
Db ||||| : : : : :  
2287 CCTAATCAATACAGA-----GAGCTTCTCATCCAAITTGG 2322  
Qy 793 SerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAspArgTyrAlaGlu 812  
Db ||||| : : : : :  
2323 CCATCTTCC-----ATGATACCACCAACAGAGTCAATACAGGCCA 2361  
Qy 813 ArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGluAla 832  
Db ||||| : : : : :  
2362 GTTCTTTAAATATTAAATCAGTCACAAAT---CCGGTACGCTTTCA-----CAGGCA 2412  
Qy 833 SerLysLeuCysAspArgArgAlaGlyGlnValValLeuTyrProLysGlu----- 850  
Db ||||| : : : : :  
2413 TCTAACCAATGAAATAATACGTGAACCTCAACTTTGTTGCTGCCAACGGGAAGCAAAATGT 2472  
Qy 850 ----- 850  
Db 2473 GGGCCTAATCCAGAAATTTCATTGCTGCTCAAAACATGCTGCGGGTTAGTAGTAGTAGT 2532  
Qy 851 -----SerMetProAlaThrHisLysLeuArg 859  
Db ||||| : : : : :  
2533 AGTAGGCCAATAGATAACTTTTCTAGTAGAGCTCTATACCGGCATTTGCTACTCTACG 2592  
Qy 860 MetMetAspProSerThrLeuAlaSerPhePro-----AsnTyrGlyThrSer----- 875  
Db ||||| : : : : :  
2593 CTCTGTGATCTCGCTGAGGTCAACACTCCGCTGACCAACGGAACACTAAATTT 2652  
Qy 876 SerArgAsnGlnMetGluSerGlnLeuHisAsnSerGlnTyrAlaHisAsnGlnTyrLys 895  
Db ||||| : : : : :  
2653 ACTAAAGACATTTTCGGCAGCCACCACTGCTCAAGAGCTTATAGAGTCTTCAACAGGG 2712





Db	6174	AGAACTCGAGGTAGATATTGTTGTAGCTTCAGAAAAACAACACCAATACAAACAGATCGTCT	62333
Qy	534	HisSerAlaIaSerThrIrsTyGlyGlySerThrArgAsnGlyGlnAsnIleHis	553
Db	6234	TTCTCGAACTTGAAAGCTAAGATACCCCTTCCTTCTACTGAAGTTCGGGATTATCTCGG	6293
Qy	554	ValLeuSerAlaGluAspGln-CysGlnMetGluThrGluAsnSerValLeuSerHis	573
Db	6294	GTGCTGCAAAAGGTACAAGCTTTGCATCTATTCTTCAATGCCATGCTTTCTCATTC	6353
Qy	573	x-----	573
Db	6354	TGTTTTCCAACTTGCACCTTATCTTGTTATATTCCAACCTGGTGTGCTTCAGTGATTGGAT	6413
Qy	574	-----Ala-LysValSerProAlaGluHisAspIleG	584
Db	6414	CAGGATGATAACGGGATTGATTACTTGGCAGGATGCTTCTGGTGAGATAGAAAGGGGA	6473
Qy	584	lnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysGlnLysLeuG	604
Db	6474	AGACTGTTATGTCCTCAAGAACATCATGTGGAGCACCAAGAGCCAAAGTCACGATGAAGAGG	6533
Qy	604	luValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuL	624
Db	6534	AGACTACGACTGAAGAGCAAAACAAC--GATGATATTCGAATGGAGATAGTCGAGCTCA	6590
Qy	624	euaIalysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnA	644
Db	6591	TGGCCAAAAACCAAGTACGAGAGGTGCTTCCCGACAAGAAGAA--GATGAAGAACA	6647
Qy	644	rgIleGlnSerLysThrAla-----AspAspAspCysValIleValAlaAlaLysA	662
Db	6648	AACAGCATCCACAAGAAACAGCACACAATCCAAAGATGCTCTACTGATTGATCTCAATG	6707
Qy	662	spGlySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGln-----LysSerL	680
Db	6708	AAACCTACGATACGGGATCTCATCTTGAGACAAACAACATCAAGACCAACCAACCGT	6767
Qy	680	euaIaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGluS	700
Db	6768	GTAGTAGCAACGCAAGGAGGAGAA-----CATTTTCTATGGGAAGACACGACA	6818
Qy	700	erProHis-----ProGlnAsn-----PheG	707
Db	6819	ACTCTCATGACTTCTTCCCAATAAGTCAGCCCTTATGTGCTCTTCTCGCTTTGGGATCTTTC	6878
Qy	707	lnSerThrGlnGlnGlnThrHisLeuArgMetGluLeuMetValThrIleAlaAlas	727
Db	6879	CTCTTACCCCAAGAAAACCGA-----GCCA	6902
Qy	727	erSerProLeuPheSerHisAspAspGlnThrIleAlaGluAlaProThrGluHisT	747
Db	6903	GCTCCATCCGGTTTTCTGGTCACAACATGTCAAGTGGCTTGGGAATTTGCCAACTGTG----	6958
Qy	747	rpGlyArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnS	767
Db	6959	--GGTAATCAGAACCCTCTCCACATCTCAFTTTCGGGTATTCAGTGTCT-----	7003
Qy	767	erProAlaAlaThrCysGlyAla-----GlnPheArgProGlyIleGlnAlaValA	784
Db	7004	-----TGTGATACTTCCAGAGGTGTTCTTAATCAATACAGA-----	7039
Qy	784	spLeuThrSerThrHisValMetGlySerSerSerAsnTyrAlaSerArgGlnProValI	804
Db	7040	-----GAAGCTTCTCATCCAATTTGGCCATCTTCC-----ATGA	7073
Qy	804	leAlaProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAlaLeArgAsnPheP	824
Db	7074	TACCACACAGAGTCAATACAGCCAGTTCCTTAATATTAAATTAATCACTCAACAAAT--C	7130
Qy	824	roSerThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnV	844

transgenic plant; flowering control; reproductive trait control.	
Arabidopsis thaliana.	
Key promoter	Location/Qualifiers
exon	1..3201
	/tag= a
	3202..3265
	/tag= c
intron	/number= 1
	3266..4159
	/tag= d
	/number= 1
exon	4160..4335
	/tag= e
	/number= 2
CDS	4241..8031
	/tag= b
intron	/product= "Arabidopsis thaliana EMP1 protein"
	4336..4447
	/tag= f
	/number= 2
exon	4448..4623
	/tag= g
	/number= 3
intron	4624..4703
	/tag= h
	/number= 3
exon	4704..4823
	/tag= i
	/number= 4
intron	4824..4902
	/tag= j
	/number= 4
exon	4903..4956
	/tag= k
	/number= 5
intron	4957..5045
	/tag= l
	/number= 5
exon	5046..6307
	/tag= m
	/number= 6
intron	6308..6447
	/tag= n
	/number= 6
exon	6448..8065
	/tag= o
	/number= 7
intron	8066..8300
	/tag= p
	/number= 7
exon	8301..8648
	/tag= q
	/number= 8
US6376751-B1.	
23-APR-2002.	
08-OCT-1999; 99US-00415946.	
09-OCT-1998; 98US-00169696.	
(REGC ) UNIV CALIFORNIA.	
Sung ZR, Aubert D, Chen L;	
WPI; 2002-453154/48.	
P-PSDB; AAO15030.	
A polynucleotide which hybridizes to an embryonic flower 1 protein, promotes early transition from vegetative to reproductive state	

Arabidopsis thaliana embryonic flower 1 (EMF1) genomic sequence.

16-AUG-2002 (first entry)

Arabidopsis thaliana embryonic flower 1 (EMF1) genomic sequence.

1

Embryonic flower 1; EMF1; gene; ds; early reproductive state tra

plant is useful to control flowering in transgenic plants.

Claim 3; Col 35-48; 30pp; English.

The invention comprises the amino acid and coding sequence of the Arabidopsis thaliana embryonic flower 1 (EMF1) protein. The EMF1 nucleotide of the invention promotes early transition from a vegetative to a reproductive state when operably linked to a plant promoter and introduced into a plant. The EMF1 nucleotide promotes early transition to a reproductive state by hybridizing to the EMF1 gene that is naturally present within the plant. The EMF1 DNA sequence is useful for controlling flowering and other reproductive traits in plants. The present DNA sequence represents the Arabidopsis thaliana embryonic flower 1 (EMF1) genomic sequence

Sequence 8648 BP; 2748 A; 1563 C; 1631 G; 2706 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	1,11e-11	Length:	8648
Score:	298.00	Matches:	225
Percent Similarity:	35.53%	Conservative:	152
Best Local Similarity:	21.21%	Mismatches:	360
Query Match:	5.39%	Indels:	324
DB:	6	Gaps:	50

US-09-721-114-2 (1-1057) x AAL43177 (1-8648)

Qy	198	AlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValala	217
Db	5090	TCTAATGTAGGACCAAGAGCTTGTGAGTCGGAGCAG-----GTAGGA	5140
Qy	218	AlaAsnValSerGluAsp-----AsnThrSerValAspValGlyAlaLeuProGlu	234
Db	5141	AACACACAGATCTAAAGAAAACTAAACAAACCATCGATGGATATTAGCAGCTGGAAGAG	5200
Qy	235	ValProGlnIleThrTrpHisIleGluValAsnGlyAla-----	247
Db	5201	AAACAAATGGGATCGGCTGTGACACAGCTTGGCTCATCTGAAATGGTGGTGGT	5260
Qy	248	---AspGlnProSerThrProLysLeuSerGluValValLeuLysArgAsnGluAsp	266
Db	5261	GAGGATACACCCCTAAGGCAACCAAGCAATCAAGGCAATTCGGCTCTGTGGAATGC	5320
Qy	267	GluAsnGlyLysThrGluGluThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro	286
Db	5321	GATACGGGTCTCAGAAAGATTAATCTTGCT-----	5353
Qy	287	AsnProMetSerGlyLysGluArg-----AspGlnValAlaGluGlnCysAsnLeu	303
Db	5354	-----ATGAGTGGGTTCAGCGGTAGGAAATCTCGCAAGGTTCGTCTACTCAGTGAGTTG	5407
Qy	304	ThrLysAspProLysProValSerGlyGlnLysCysGluGlnIleCysAsnGluProCys	323
Db	5408	CTTGGTAATACAAAACCAAGTGGTGTAGTAACATCAGAAAA-----	5449
Qy	324	GluGluValValLeuLysArgSerSerLysSerLysArgLysThrAspLysLysLeuMet	343
Db	5450	GAACAGTCTGCTTGAAGAGGATCAGTAGAGGTGMAAA-----AGAAAGTGTGTA	5503
Qy	344	LysLysGlnGlnHis-----SerLysLysArgThrAlaGlnAlaAspValSerAsp	360
Db	5504	CCTGAAACAAATATTATGTCGCGGATATTGAGTCAATGGGTGCAACCTCTGAAAAATGCT	5563
Qy	361	AlaLysLeuCys-----	364
Db	5564	TCCAAAGTGTGACTCTGATCAAGGTAATAGTCAATCACTGATAGTGGTTGACAGA	5623
Qy	365	-----ArgArgLysProLysValArgLysValArgLeuLeuSerGluIleAsn	379
Db	5624	ACTCCATTTAAGGGCTAAGCAGAGAAACAGAGATTTTCAGGTGTTGTGACGAGTTGTACCA	5683
Qy	380	AlaAsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspPro	399
Db		-----	

Db	5684	TCACCTTCTGTGAAACTTCACAA-----GAAGGTATCAAGGAGCATGATGATGATCCT	5737
Qy	400	CysGluAspAspArgSerThr-----	408
Db	5738	AGTAAG-----AGATCAACTCTCTGGCACTCTTTATTCTACTCGAAACGATTTCTGTCTCT	5791
Qy	409	ValProMetGluValSerMetAspIleProValSerAsnHisThrValGlyLysAspGly	428
Db	5792	TGTCT-----COGGTACTCAGAGAAC-----CAGAGGAAG	5824
Qy	429	LeuLysSerSerLysAsnLysThrLysArgLysTyrSerAspValValAspAspGlySer	448
Db	5825	CTCAGTTTATCCCAAGAGAACAAAG-----CCTGTAATCGATATATGGAAG	5875
Qy	449	SerLeuMet-----AsnTrpLeuAsnGlyLysLys-----LysArgThrGly	462
Db	5876	AGCACTGTGATCAGTTTGTAGTAAAGCATGTGAGGAGTCAAGTAACTGCGATACCTGT	5935
Qy	463	SerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysValThrPro	482
Db	5936	CTTTCCATGAACACAGTATCCCAACTCGAGACTTATTGAATCGGAAGGGTGGGGCGT	5995
Qy	483	ThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAsp-----	499
Db	5996	TTATTTGACAAACGGTTTGGCTTCAGATGGATATTTTCAGAAATATCTCTCTCAGGTTAAT	6055
Qy	500	-----ThrAsnMetHisLysThrAspValCysGlnHisValSerGluIleSer	515
Db	6056	GATAAGCCGATACATCTTTGCAATTTGCAAGACAATGATTTATCTGAGGTCAAGACGCG	6115
Qy	516	ThrGlnArgCys-----SerSerLysGlyLysThrAlaGly---Leu	528
Db	6116	GAACCAAACTGCTTCGAGATTTTAGTTCTCTCTTAATTCAGTTCAGGTGGATGGTTG	6175
Qy	529	SerLysGly-----	533
Db	6176	AGAAGTGGAGTAGATATTTGTGACTTTCAGAAACCAACCAACATACAAACAGATGTCT	6235
Qy	534	HisSerAlaAlaSerThrLysThrGlyGluSerThrArgAsnGlyGlnAsnIleHis	553
Db	6236	TTCTGAACTTGAAGCTAAGATACCCCTCTCTTCTAGAGTTCGGATTTATCTCGG	6295
Qy	554	ValLeuSerAlaGluAspGln-CysGlnMetGluThrGluAsnSerValLeuSerHis	573
Db	6296	GTGCTCAAAAGGTACAGCTTTGCATCTATTTCTTCAATCGCCATGCTTCTCATTC	6355
Qy	573	-----	573
Db	6356	TGTTTTCAACTTGCACCTATTCTTGTATATTTCCAAACTGGTGTCTTCAGTGATGGAT	6415
Qy	574	-----Ala-LysValSerProAlaGluHisAspIleG	584
Db	6416	CAGGATGATAACGGGATTTGATTTACTTGGCAGGATCTTCTGTCAGATAGAAAGGGA	6475
Qy	584	InLeuMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysGlnLysLeuG	604
Db	6476	AGACTGTTATGTTCCAAAGAACATCATGAGCACCAAGAACCAAGTCCAGATAGAAAGG	6535
Qy	604	IuValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeu	624
Db	6536	AGACTACGACTGAAGAGCAAAACCAAC---GATGATATTTCAATGGAGATAGTGGAGCTCA	6592
Qy	624	euAlaLysAsnGlnHisGluArgGlnLeuMet---ThrGluThrAspCysSerAspIleA	643
Db	6593	TGGCCAAAACCAAGTACGAGAGGTGTCTTCCGACAAAGAGAGATGTTAGCAACAAAC	6652
Qy	643	snArgIleGlnSerLysThrThrAlaAspAspCysValIleValAlaAlaLysAspG	663
Db	6653	AGCCATCAGCAAGAAACAGACACAAATCCAAAGATGCTCTACTGATGTATCTCAATGAAA	6712
Qy	663	lySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGln-----LysSerLeuA	681
Db	6713	CCTACGATTAACGGATCTCATTGAGGACAAACACATCAAGACCACCAAAACCGTGTA	6772

QY 681 laSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGluSerP 701  
Db 6773 GTAGCAACGAGGAGGAGAA-----CATTTCTATATGGAAGACGACGACT 6823  
QY 701 roHis-----ProGlnAsn-----PheGlns 708  
Db 6824 CTATGACTCTTCCCAATAAGTCAGCTTATGTCCTCTCCGTTGGATCTTTCTCTC 6883  
QY 708 erThrGlnGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerS 728  
Db 6884 CTACCCCAAGAAACCGA-----GCCAGCT 6907  
QY 728 exProLeuPheSerHisAspGlnTyrlleAlaGluAlaProThrGluHisTrpG 748  
Db 6908 CCATCCGTTTCTGCTCACAACCTGTCAGTGGCTTGGGAATTTGCCAAGCTGTC-----G 6961  
QY 748 lyArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSerP 768  
Db 6962 GTAATCAGAACCCCTTCTCCATCCTCATTTCCGGTATTACGTCT----- 7005  
QY 768 roAlaAlaThrCysGlyAla-----GlnPheArgProGlyIleGlnAlaValAspL 785  
Db 7006 --TGTGATCTTGGCAGAGTCTTCTTAATCATACAGA----- 7041  
QY 785 euThrSerThrHisValMetGlySerSerSerAsnTyrlleAlaSerArgGlnProValIleA 805  
Db 7042 --GAAGCTTCTCATCCCAATTTGGCCATCTTCC-----ATGATAC 7078  
QY 805 laproLeuAspArgTyrlleAlaGluAlaValHisAlaArgAsnPheProS 825  
Db 7079 CACCACAGAGTCATACAGCAGCTTCTTTAATATATATCATCTCAACAAT-----CCGG 7135  
QY 825 erThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValV 845  
Db 7136 GTAGCTTTCA-----CAGCACTTAACAATGAATACATGGAACCTCAACTTTCTTG 7189  
QY 845 alLeuTyroLysGlu----- 850  
Db 7190 CTGCCAAGCGGAGCAAAATGTCGGCTAATCAGAAATTTTCATTTGGTGTCAACATG 7249  
QY 851 -----SerM 852  
Db 7250 CTGCTGGGTTAGTAGTAGTAGTAGTAGGCAATAGATAACTTTTCTAGTGAGAGCTCTA 7309  
QY 852 etProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPhePro----- 870  
Db 7310 TACCGCATTCATCTACTCAGCCTTCTGGATCTCGCTGAGGTCAACGACTCCCGCTG 7369  
QY 871 --AsnTyroLysThrSer-----SerArgAsnGlnMetGluSerGlnLeuHisAsnSerG 888  
Db 7370 ACCAACCGGAACACTAAATTTACTAAAGACATTTCCGCCAGCCACCGTCCCAAG 7429  
QY 888 lntyrAlaHisAsnGlnTyrllysGlySerThrSerThrSerThrGlySerAsnLeuAsnG 908  
Db 7430 AGTTTATAGACTTCAACACAGGGGACTCTAGTAAGTCAGCCTACTCAACTAAG----- 7482  
QY 908 lyLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgP 928  
Db 7483 --CAGATACCT-----TTTGATCTATACAGAGAGATTACACAGAGCCTTCCCGGA 7534  
QY 928 roLeuArgPro-----HisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIleA 947  
Db 7535 AGAGTTTCCCATCACTCCACTATTTGGAGC-----TCTTCACTTTCAATTTCAAAATG 7588  
QY 947 laAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrllysLeuGlyValSerThrGlyI 967  
Db 7589 CTTTCATGGAGTCTCAT----- 7605  
QY 967 leThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheS 987  
Db 7606 -----CATCAGAGAGAGAAACCAAGAGAGACACC-----TTTG 7642

QY 987 erAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAsp----- 1003  
Db 7643 CTCCTGTTTACAACTACTCATGAAAGACCGGTGTTTGCAGCAGCAACGACCAAGCGAAGT 7702  
QY 1004 -----PheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysM 1022  
Db 7703 TCCAGCTCTGGAGACATCGAATTC-----A 7729  
QY 1022 etValHisProLeu-----AspArgPheValArgGlnAsp----- 1033  
Db 7730 TGATCTTCTTCTTGAATTTTCATGACGGATAAAGAAAAGAAAACAAAAGAGAAAAGCAG 7789  
QY 1034 ----- 1035  
Db 7790 AGAGCTGCATTAACAATGCCTCTGGGGACCTGTGAAGAACAGTTCCTGACCCATTTGTT 7849  
QY 1035 ysileThrAsnLysAsnProAlaAspPheThrThrIleSerAsnAspAsnGluTyrlMet 1054  
Db 7850 GCAGGCTCATAGAAACCTCTGTGATTCACCATTCCTGACCTGGGAATGTTTACATG 7908  
RESULT 7  
ABQ77909  
ID ABQ77909 standard; DNA; 8648 BP.  
XX ABQ77909;  
AC ABQ77909;  
XX 24-JAN-2003 (first entry)  
DT Arabidopsis thaliana EMFL gene.  
DE Arabidopsis thaliana EMFL gene.  
XX EMFL; embryonic flower 1; rice OsEMFL1 homologue; plant; floral repressor;  
KW reproductive development; flower development; transgenic plant;  
KW antisense suppression; transgenic; reciprocal negative interaction;  
KW flower meristem identity gene; flowering time; shoot development;  
KW seed yield; agriculture; gene; ds.  
XX Arabidopsis thaliana.  
XX Key Location/Qualifiers  
PH 4241..8031  
FT CDS /tag= a "EMFL protein"  
FT FT 4241..4335  
FT FT /product= "EMFL protein"  
FT FT /tag= b  
FT FT /number= 1  
FT FT 4335..4447  
FT FT /tag= c  
FT FT /tag= c  
FT FT /number= 1  
FT FT 4448..4623  
FT FT /tag= d  
FT FT /number= 2  
FT FT 4624..4703  
FT FT /tag= e  
FT FT /number= 2  
FT FT 4704..4823  
FT FT /tag= f  
FT FT /number= 3  
FT FT 4824..4902  
FT FT /tag= g  
FT FT /number= 3  
FT FT 4903..4956  
FT FT /tag= h  
FT FT /number= 4  
FT FT 4957..5045  
FT FT /tag= i  
FT FT /number= 1  
FT FT 5046..6307  
FT FT /tag= j  
FT FT /number= 5  
FT FT 6308..6447  
FT FT /tag= k  
FT FT /number= 5  
FT FT 6448..8031  
FT FT exon

```

FT      /*tag= 1
FT      /number= 6
PN      WO200280659-A1.
PD      17-OCT-2002.
PF      05-APR-2002; 2002WO-US012675.
PR      06-APR-2001; 2001US-00828068.
XX      (REGC ) UNIV CALIFORNIA.
PI      Moon Y, Chen L, Sung ZR;
XX      WPI; 2003-046831/04.
DR      P-PSDB; ABB99878.
XX      New isolated OsEMF1 nucleic acid molecule and encoded polypeptide, useful
PT      for plant genetic engineering, in particular controlling reproductive
PT      development in rice.
XX      Example 2; Page; 47pp; English.
XX      The invention relates to a rice EMF (embryonic flower) gene designated
CC      OsEMF1 (CDNA given in ABQ77908) and its encoded protein (ABB99878). The
CC      OsEMF1 protein has a molecular weight of 116.4 kD and exhibits 37%
CC      homology and 20% identity with Arabidopsis thaliana EMF1 protein
CC      (ABB99879). OsEMF1, like other EMF gene products, acts as a floral
CC      repressor, suppressing the transition from vegetative growth to
CC      reproductive development. It also delays the inflorescence to flower
CC      transition, indicating that there is a reciprocal negative interaction
CC      between OsEMF and flower meristem identity genes. The invention also
CC      encompasses a transgenic plant comprising an expression cassette
CC      containing an OsEMF1 nucleic acid (particularly in an antisense
CC      orientation) under the control of a plant promoter. OsEMF1 nucleic acids
CC      and proteins may be used for controlling reproductive development in
CC      plants, particularly monocotyledonous plants and especially rice. In
CC      particular, OsEMF1 nucleic acids and proteins may be used to control
CC      flowering time, shoot development and seed yield. Controlling or
CC      inhibiting the expression of genes which mediate these processes enables
CC      new varieties of rice with different flowering times and seed yield to be
CC      developed. The present sequence represents the Arabidopsis thaliana EMF1
CC      gene which is referred to in an exemplification of the invention. Note:
CC      The present sequence is not given in the specification, but was obtained
CC      from GenBank (accession number AP319968)
XX
SQ      Sequence 8648 BP; 2748 A; 1563 C; 1631 G; 2706 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.11e-11 Length: 8648
Score: 298.00 Matches: 225
Percent Similarity: 35.53% Conservative: 152
Best Local Similarity: 21.21% Mismatches: 360
Query Match: 5.39% Indels: 324
DB: Gaps: 50

US-09-721-114-2 (1-1057) x ABQ77909 (1-8648)

Qy      198 AlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyAspValala 217
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      5090 TCTAATGTTAGGACCAAGCAGCAAGAGCTTGAGTCCGGAGCAG-----GTAGGA 5140
      ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy      218 AlaAsnValSerGluuAsp-----AsnThrSerValAspValGlyAlaLeuProGlu 234
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      5141 AACACACAGATCTAAAGAAAAAATAACAAACACCCGATGGATATTAGCAGCTGGAAGAG 5200
      ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy      235 ValProGlnIleThrTrpHisIleGluValAsnGlyAla----- 247
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db      5201 AAACAAATGTGGATCAGGCTGTGCACAGCTTCGGCTCATCTGAAATTCTGGTGTGTT 5260
      ::::: ::::: ::::: ::::: ::::: ::::: :::::

Qy      248 ---AspGlnProProSerThrProLysLeuSerGluuValValLeuLysArgAsnGluuAsp 266
      ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

QY 554 ValIeuSerAlaGluAspGln-CysGlnMetGluThrGluAsnSerValIeuSerHisSe 573  
Db 6296 GTGCTGCAAAAGGTACAAAGCTTTGCTATATTTTCTCAATGCCAGTCTTCTCAAT 6355  
QY 573 T----- 573  
Db 6356 TGTTTTCCAACTGCACCTATTCTTGATATTCCAAAGTGGTGCTTCAGTGATTCGAT 6415  
QY 574 -----Ala-LysValSerProAlaGluHisAspIleG 584  
Db 6416 CAGGATGATACGGGATTTACTTTGCGAGGATGCTTCTGTCAGATGAGAAGGGGA 6475  
QY 584 InIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysGlnLysLeuG 604  
Db 6476 AGACTGTTATGGTCCAGAACATCATGAGCACCAAGCAAGCAAGTCCAGATAGAAAGG 6535  
QY 604 luValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuL 624  
Db 6536 AGACTTACGACTGAAGCAAAACAAAC---GATGATATTTCCAATGGAGATAGTGAGCTCA 6592  
QY 624 euAlaLysAsnGlnHisGluArgGlnLeuMet---ThrGluThrAspCysSerAspIleA 643  
Db 6593 TGGCCAAACCACTACGAGAGGTCTTCCCGACAAGAGAGAGATGTTAGCAACAAC 6652  
QY 643 snArgIleGlnSerLysThrAlaAspAspAspCysValIleValAlaLysAspG 663  
Db 6653 AGCCATCACAGAAACACGACACAAATCCAAAGATGCTTACTGATTCATCAATGAA 6712  
QY 663 LysSerAspTyrAlaSerSerValPheAspThrAsnSerGlnGln-----LysSerLeuA 681  
Db 6713 CCTAGATAACGGGATCTCACTGAGGACAAACACATCAAGACCAACCAACCGGTGA 6772  
QY 681 laSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnLysSerP 701  
Db 6773 GTAGCAACGCAAGGAGGCAAGAA-----CATTTCTCTATCGGAAGACAGCAGAACT 6823  
QY 701 roHis-----ProGlnAsn-----PheGlns 708  
Db 6824 CTCATGACTTCTCCCAATAGTCAGCCTTATGTCCTTCTCCCTTTGGGATCTTCTCTC 6883  
QY 708 erThrGlnGluGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaLysSerS 728  
Db 6884 CTACCCAGAAACCGA-----GCCAGCT 6907  
QY 728 erProLeuPheSerHisHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTrpG 748  
Db 6908 CCATCCGGTCTTCTGTCACAACTGTCAGTGGCTTGGGAATTTCCCAACTGTG-----G 6961  
QY 748 lyArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSerP 768  
Db 6962 GTATCAGAACCTTCTCCATCTCTCAATTCGGGTATTACGGCT----- 7005  
QY 768 roAlaAlaThrCysGlyAla-----GlnPheArgProGlyIleGlnAlaValAspL 785  
Db 7006 --TGTGATCTTGCAGAGTGTCTTAATCAATACAGA----- 7041  
QY 785 euThrSerThrHisValMetGlySerSerAsnTyrAlaSerArgGlnProValIleA 805  
Db 7042 --GAAGCTTCTCAATCCAAATTTGGCCACTTCC-----ATGATAC 7078  
QY 805 laProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProS 825  
Db 7079 CACCACAGTCAATACAAAGCCAGTTCTTTAAATATTAATCAGTCAACAAT---CCGG 7135  
QY 825 erThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValV 845  
Db 7136 GTACGCTTTCA-----CAGGCATCTAACAAATGAAATACATGGAACCTCACTTTGTG 7189  
QY 845 alLeuTyrProLysGlu----- 850  
Db 7190 CTGCCAACGGAGCAAAAATGCTGGCCTAATCCAGAAATTTTCATTTTCGCTGCAACATG 7249

851 -----SerM 852  
Db 7250 CTGCTGGGTTAGTAGTAGTAGTAGCCCATATTAACATTTTCTAGTGAGAGCTCTA 7309  
QY 852 etProIaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPhePro---- 870  
Db 7310 TACCGCATTTGATCTACTACTCAGCTTCTGGATCTCTGCTGAGGTCAACAGCTCCGCTG 7369  
QY 871 --AsnTyrGlyThrSer-----SerArgAsnGlnMetGluSerGlnLeuHisAsnSerG 888  
Db 7370 ACCAACACGGAAACACTAATTTACTAAAGACATTTTCGGCCAGCAACCACTCCCAAG 7429  
QY 888 InTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerThrSerThrSerThrSerThr 908  
Db 7430 AGTTTATAGAGCTTCAACACAGGGGACTCTAGTAAGTCAGCTACTCAACTAAG----- 7482  
QY 908 LysLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgP 928  
Db 7483 --CAGATACT-----TTTGATCTATACAGCAAGATTCACACAGAGCTTCCCGGA 7534  
QY 928 roLeuArgPro---HisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIleA 947  
Db 7535 AGAGTTTCCCATCACTCCACCTATTGGGACG-----TCTTCACTTTCAATAAG 7588  
QY 947 laAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSerThrGlyI 967  
Db 7589 CTTATGGAGTCTCTCAT----- 7605  
QY 967 leThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheS 987  
Db 7606 -----CATCAGAGAGAAACCAAGAGAAAGACACC-----TTTG 7642  
QY 987 erAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAsp----- 1003  
Db 7643 CTCTCTTTTACAACTACTCATGAAAGCCGGTGTGTTGCAAGCAGCAACGACCAAGCAAGT 7702  
QY 1004 -----PheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysM 1022  
Db 7703 TCAGCTGTGGGAGCATCGAATTC-----A 7729  
QY 1022 etValHisProLeu-----AspArgPheValArgGlnAsp----- 1033  
Db 7730 TGATGCTTCTTTGAAATTTTCATGACGGATAAAGAAAGAAAGAAAGAGAGAGAGCAG 7789  
QY 1034 -----IleC 1035  
Db 7790 AGAGCTGCAATACAAATGCTCTGCGGACCTGTGAGAACAGTCTTGAGCCCATTTGT 7849  
QY 1035 ysIleThrAsnLysAsnProAlaAspPheThrThrIleSerAsnAspAsnGluTyrMet 1054  
Db 7850 GCAGCGTCAATAGAAACCTGCTGATTTCCCACTTCTCGAACCTCGGAATGTTTACATG 7908  
RESULT 8  
AA14872 standard; DNA; 17341 BP.  
ID AAA14872  
XX AC AAA14872;  
XX DT 08-AUG-2000 (first entry)  
XX DX Genomic DNA sequence from Arabidopsis thaliana CD82 clone.  
XX DB Embryonic flower gene-1; EMF-1; reproductive development; flowering;  
XX KW early flowering; uniform flowering; ss.  
XX OS Arabidopsis thaliana.  
XX PN WO200022132-A2.  
XX PD 20-APR-2000.  
XX PF 08-OCT-1999; 99WO-US023543.  
XX

PR 09-OCT-1998; 98US-00169696.  
 XX (REGC ) UNIV CALIFORNIA.  
 XX PA  
 XX PI Sung ZR, Aubert D, Chen L;  
 XX WPI; 2000-317987/27.  
 DR An isolated nucleic acid molecule comprising embryonic flower genes is  
 XX useful for modulating reproductive development in plants.  
 XX Claim 3; Page 30-33; 38pp; English.  
 XX The present sequence represents the genomic sequence of the Arabidopsis  
 CC thaliana genomic DNA from CD82 clone. The sequence comprises the  
 CC embryonic flower gene-1 (EMF-1). The EMF genes play an important role in  
 CC reproductive development in plants. Control of the expression of the  
 CC genes is therefore useful in controlling flowering and other functions in  
 CC plants. The EMF-1 polynucleotides are useful for modulating reproductive  
 CC development in plants. For example, enhanced expression of EMF-1 is  
 CC useful to prevent flowering in grasses or vegetable crops. Controlling or  
 CC inhibiting expression of the genes is useful to promote early flowering  
 CC in forest trees and agronomical crops or to ensure uniform flowering in  
 CC various agronomic crops or ornamentals.  
 XX  
 SQ Sequence 17341 BP; 5468 A; 3114 C; 3409 G; 5350 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 2.78e-11 Length: 17341  
 Score: 298.00 Matches: 225  
 Percent Similarity: 35.53% Conservative: 152  
 Best Local Similarity: 21.21% Mismatches: 360  
 Query Match: 5.39% Indels: 324  
 DB: 3 Gaps: 50

US-09-721-114-2 (1-1057) x AAA14872 (1-17341)

QY 198 AlaAsnThrAspSerProMetIysAspLeuGlnGlyProAlaGlnAsnTyrAspValala 217  
 DB TCTATGTTAGGACGACGAGCAAGCTTGTGAGTCCGGAGCA-----GTAGGA 5140  
 QY 218 AlaAsnValSerGluAsp-----AsnThrSerValAspValGlyAlaLeuProGlu 234  
 DB TCTATGTTAGGACGACGAGCAAGCTTGTGAGTCCGGAGCA-----GTAGGA 5140  
 QY 5141 AACACACAGATCTAAAGAAAAAATAAACAAACCATCGATGATATTAGCAGCTGGAAGAG 5200  
 QY 235 ValProGlnIleThrTphIleGluValAsnGlyAla----- 247  
 DB 5201 AAACAAATGTGGATCAGGCTGTGACACAGTTCGGCTCATCTGAAATGTGGTGGTT 5260  
 QY 248 ---AspGlnProProSerThrProIysLeuSerGluValValLeuLysAsnGluAsp 266  
 DB GAGGATACACACCTAAGCAACCAAGAAATCAATAAGGCATTCCGGCTCGATGGAATGC 5320  
 QY 267 GluAsnGlyLysThrGluGluThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro 286  
 DB GATAACGGGTCTACAGAAAGTATAAATCTGCT----- 5353  
 QY 287 AsnProMetSerGlyLysGluArg-----AspGlnValAlaGluGlnCysAsnLeu 303  
 DB ATGAGTGGGTTCGACGCTAGGAAATCTCGACAGTTCGCTACTACTAGTGGTTG 5407  
 QY 304 ThrLysAspProLysProValSerGlyGlnLysCysGluGlnIleCysAsnGluProCys 323  
 DB CTGTGTAATACAAAACCAAGCTGTGTAGTACATCAGAAAA----- 5449  
 QY 324 GluGluValValLeuValAsnSerLysSerLysAspValSerLysThrAspLysValLeuMet 343  
 DB GAAGAGTCTGCTTTGAAGAGGAATCAGTTAGAGTTCGAAAA-----AGAAAGTTGTTA 5503  
 QY 344 LysLysGlnGlnHis---SerLysLysArgThrAlaGlnAlaAspValSerAsp----- 360  
 DB CCTGAAACAAATATGTCAGCGCGGATATTGAGTACATGAGGTGCACCTCTGAAATGCT 5563

QY 361 AlaLysLeuCys----- 364  
 DB TCCAAAAGTTGTGACTCTGTGATCAAGGTAAATAGTGAATCAACTGATAGTGGTTTGACAGA 5623  
 QY 365 -----ArgArgLysProLysLysValArgLeuLeuSerGluLeuLeuLeu 379  
 DB ACTCCATTAAAGGTAAAGCAGAAAGCAAGATTTCGGTGTGTGTGACGAGTTGTGTACCA 5683  
 QY 380 AlaAsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspPro 399  
 DB TCACTTCCTTGTGAACACTTCACAA-----GAAGGTATCAAGGAGCATGATGACAGATCT 5737  
 QY 400 CysGluAspAspArgSerThr-----IlePro 408  
 DB AGTAAG-----AGATCAACTCTCGGCACCTCTTTATTCATCTGGAACAGATCTGTTCCT 5791  
 QY 409 ValProMetGluValSerMetAspIleProValSerAsnHisThrValGlyGluAspGly 428  
 DB TGTCTCT-----COGGTACTCAGAGAAC-----GAGAGGAAG 5824  
 QY 429 LeuLysSerSerLysAsnLysThrLysArgLysTyrSerAspValValAspAspGlySer 448  
 DB CTTCATGTTTACCCAGAGAAAGACAAAGAA-----CCTGTATCGATATATGGAAG 5875  
 QY 449 SerLeuMet-----AsnTyrLeuAsnGlyLysLys-----LysArgThrGly 462  
 DB AGACTGTGTGATCAGTTTGTAGTAAGCGCATGTGATGAGAGTCAAGTTAACTCGCATCTGCT 5935  
 QY 463 SerValHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysValThrPro 482  
 DB CCTTCATGAACACACAGTATCCCAACTTCGAGACTTATTGAATGGGAAAGGTGGCGCGT 5995  
 QY 483 ThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAsp----- 499  
 DB TTATTTGACAAACCGTTGGCTTCAGATGGATATTTTCAGAAAAATATCTCTCAGGTATAT 6055  
 QY 500 -----ThrAsnMetHisLysThrAspValCysGlnHisValSerGluIleSer 515  
 DB GATAAGCGGATAACATCTTTGCAATTTGCAAGACAATGATTATGTGAGTCAAGACAGCGG 6115  
 QY 516 ThrGlnArgCys-----SerSerLysGlyLysThrAlaGly---Leu 528  
 DB GAACCAAACTGTCTTCGAGATTTTAGTTCCTCTCTTCTAATCCAGCTCAGGTGGATGTTG 6175  
 QY 529 SerLysGly-----LysThr 533  
 DB AGACTGGAGTAGATATTGTTGATTCAGAAACACCAACCAACATCAACACAGATCTGCT 6235  
 QY 534 HisSerAlaAlaSerThrLysTyrGlyGlyGlySerThrArgAsnGlyGlnAsnIleHis 553  
 DB TTCTCGAACTTGAAGCTAAGATACCCCTTCTTCTACTGAACTTCGGGATTTATCTCGG 6295  
 QY 554 ValLeuSerAlaGluAspGln-CysGlnMetGluThrGluAsnSerValLeuSerHis 573  
 DB GTGCTGCAAAAGGTACAGCTTTGCATCTATTTCTTCATCGCCATGCTTCTTCATTC 6355  
 QY 573 r----- 573  
 DB TGTTCACACTTGCACCTATTCTCTGTATATTCACAACTGGTGTGCTTCAGTGAITGGAT 6415  
 QY 574 -----Ala-LysValSerProAlaGluHisAspIleG 584  
 DB CAGGATGATAACCGGATTTGATTTACTTGGCAGGATGCTTCTGTGTCGATAGAAAGGGA 6475  
 QY 584 InIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysLysLysLysLeuG 604  
 DB AGACTGTATGTCCTCAAGAACATCATGAGCACCACCAAGAGCCAAAGTCACGATAGAAAGG 6535  
 QY 604 LuValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeu 624  
 DB AGACTACGACTGAAGAGCAAAACAAAC---GATGATATTCATGAGATGATGAGATCTCA 6592  
 QY 624 euAlaLysAsnGlnHisGluArgGlnLeuMet---ThrGluThrAspCysSerAspIleA 643



Db 6593 TGCCCAAAACACAGTACGAGAGTGCTCTCCCGACAAAGAGAGATGTTAGCAAAAC 6652  
Qy 643 snargileGlnSerLysThrAlaAspAspCysValIleValAlaLysAspG 663  
Db 6653 AGCCATCACAGAAACAGCAGACAAATCCAAAGATGCTCTACTGATTGATCTCAATGAA 6712  
Qy 663 lysSerAspTyrAlaSerSerValPheAspThrAsnSerGlnGln-----LysSerLeuA 681  
Db 6713 CTACCATTAACGGGATCTCACTTGAGGACAAACACATCAAGACCAACCAACCGTGT 6772  
Qy 681 lAserGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGluSerP 701  
Db 6773 GTAGCAACCGCAAGGAGAGAA-----CATTTTCTATGGAAGACAGCAGAACT 6823  
Qy 701 rohis-----ProGlnAsn-----PheGlnS 708  
Db 6824 CTCATGACTCTTCCCAATAAGTCACGCTTATGTCCTTCCGTTGGGATCTTCCCTC 6883  
Qy 708 erThrGlnGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerS 728  
Db 6884 CTACCCCAAGAAACCGA-----GCCAGCT 6907  
Qy 728 erProLeuPheSerHisHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTrpG 748  
Db 6908 CCATCCGGTCTTCTGCTCACAACCTGTCAGTGGCTTGGGAATTTGCCAACTGTG-----G 6961  
Qy 748 lVArgLysAspAlaLysLysLeuThrTrpGlnGlnPheLysAlaThrThrArgAsnSerP 768  
Db 6962 GTAATCAGAACCTTCTCCATCTCTCAATTCGGGTATACGTGCT----- 7005  
Qy 768 roAlaAlaThrCysGlyAla-----GlnPheArgProGlyIleGlnAlaValAspL 785  
Db 7006 --TGTGATACTTCCAGAGTGTTCCTAATCAATACAGA----- 7041  
Qy 785 euThrSerThrHisValMetGlySerSerSerAsnTyrAlaSerArgGlnProValIleA 805  
Db 7042 --GAAGCTTCTCAATTCGCAATTCCTTC-----ATGATAC 7078  
Qy 805 laProLeuAspArgTyrAlaGluAlaValAsnGlnValHisAlaArgAsnPheProS 825  
Db 7079 CACCACAGTCAATACAGCCAGTCTTCTTAAATATTAATAGTCACAAAT---CCGG 7135  
Qy 825 erThrIleAlaThrMetGluAlaSerLysLeuCysAspArgAsnAlaGlyGlnValV 845  
Db 7136 GFACGCTTTCA-----CAGGCATCTAAACAATGAAATACATGGAACCTCACTTTGTG 7189  
Qy 845 allLeuTyrProLysGlu----- 850  
Db 7190 CTGCCAACGGGAGCAAAATATGCGGCTAATCCAGAAATTTTCATTTGGCTGCACAAATG 7249  
Qy 851 -----SerM 852  
Db 7250 CTGCTGGGTTAGTAGTAGTAGTAGGCGCAATAGATACTTTTCTAGTAGAGACTCTA 7309  
Qy 852 etProAlaThrHisLeuArgMetMetAspProSerThrLeuAlaSerPhePro----- 870  
Db 7310 TACGGGATTCGATCTACTCAGCTTCTGATCTCTCGCTGAGGTCAACGACTCCGCTG 7369  
Qy 871 --AsnTyrGlyThrSer-----SerArgAsnGlnMetGluSerGlnLeuHisAsnSerG 888  
Db 7370 ACCAACACGGGAAACACTAAATTTTACTAAAGACATTTTCGCCAGCAACACAGTCCAAAG 7429  
Qy 888 lntYrAlaHisAsnGlnTyrLysGlySerThrSerThrSerThrGlySerAsnLeuAsnG 908  
Db 7430 AGTTATAGACTTCAACAGGGGACTCTAGTAGTCAAGCTTACTCACTAAG----- 7482  
Qy 908 lylsilleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgP 928  
Db 7483 --CAGATACCT-----TTTGATCTATACAGCAAGAGATTTCACACAGAGCCTTCCCGGA 7534  
Qy 928 roLeuArgPro---HisProArgValGlyValLeuGlnGlySerLeuLeuGlnLysGluLea 947

Db 7535 AGAGTTTCCCATCCATCCACTATTGGGACG-----TCTTCACTTTCAATTCAAAATG 7588  
Qy 947 laAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSerThrGlyI 967  
Db 7589 CTTTCATGGAGTCTCTCAT----- 7605  
Qy 967 leThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheS 987  
Db 7606 -----CATCAGAGAGAAACCAAGAGAAAGACACC-----TTTG 7642  
Qy 987 erAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAsp----- 1003  
Db 7643 CTCCTGTTTACATACTACTGAAAGCCGGTGTTCGCAAGCAGCAACGACCAAGCGAAGT 7702  
Qy 1004 -----PheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysM 1022  
Db 7703 TCCAGCTCTCTGGAGCATCGAATTC-----A 7729  
Qy 1022 etValHisProLeu-----AspArgPheValArgGlnAsp----- 1033  
Db 7730 TGATGCTCTCTTTGAATTTTTCATGACGAGTAAAGAAAGAAACAAAGAGAAAGCAG 7789  
Qy 1034 ----- 1035  
Db 7790 AGAGCTGCATTAACAATGCTCTCTCGGGACCTGTGAAGAACAGTTCTGGACCCATTGTG 7849  
Qy 1035 ysilleThrAsnLysAsnProAlaAspPheThrThrIleSerAsnAspAsnGluTyrMet 1054  
Db 7850 GCAGCGTCAATAGAAACCTCTGCTGATTTCCAGCATTCCTGGAACCTGGAATGTTACATG 7908  
RESULT 9  
AAL43176  
ID AAL43176 standard; DNA; 17341 BP.  
XX AAL43176;  
XX 16-AUG-2002 (first entry)  
DT Arabidopsis thaliana embryonic flower 1 gene contained within a vector.  
XX Embryonic flower 1; EMF1; gene; ds; early reproductive state transition;  
XX transgenic plant; flowering control; reproductive trait control;  
XX pBluescript vector.  
OS Arabidopsis thaliana.  
OS Synthetic.  
OS Chimeric.  
FH Key Location/Qualifiers  
FT 1..3201  
FT promoter /\*tag= a  
FT CDS 4241..8031  
FT /\*tag= b  
FT /product= "Arabidopsis EMF1 protein"  
FT /note= "The CDS contains introns"  
XX US376751-B1.  
XX 23-APR-2002.  
XX 08-OCT-1999; 99US-00415946.  
XX 09-OCT-1998; 98US-00169696.  
XX (REGC ) UNIV CALIFORNIA.  
XX Sung ZR, Aubert D, Chen L;  
XX WPI; 2002-453154/48.  
XX P-PSDB; AAO15030.  
XX A polynucleotide which hybridizes to an embryonic flower 1 protein and  
XX promotes early transition from vegetative to reproductive state when in a

PT plant is useful to control flowering in transgenic plants.

XX Example 1; Col 19-34; 30pp; English.

XX The invention comprises the amino acid and coding sequence of the  
 CC Arabidopsis thaliana embryonic flower 1 (EMF1) protein. The EMF1  
 CC nucleotide of the invention promotes early transition from a vegetative  
 CC to a reproductive state when operably linked to a plant promoter and  
 CC introduced into a plant. The EMF1 nucleotide promotes early transition to  
 CC a reproductive state by hybridizing to the EMF1 gene that is naturally  
 CC present within the plant. The EMF1 DNA sequence is useful for controlling  
 CC flowering and other reproductive traits in plants. The present DNA  
 CC sequence represents the Arabidopsis thaliana embryonic flower 1 (EMF1)  
 CC gene sequence contained within a pBluescript vector

XX SQ Sequence 17341 BP; 5468 A; 3114 C; 3409 G; 5350 T; 0 U; 0 Other;

# Alignment Scores:

Pred. No.:	2,78e-11	Length:	17341
Score:	298.00	Matches:	225
Percent Similarity:	35.53%	Conservative:	152
Best Local Similarity:	21.21%	Mismatches:	360
Query Match:	5.39%	Indels:	324
DB:	6	Gaps:	50

US-09-721-114-2 (1-1057) x AAL43176 (1-17341)

Qy	198	AlaAsnThrAspSerProMetLeuAspLeuGlnGlyProAlaGlnAsnTyrAspValala	217
Db	5090	TCTAATGTTAGGCAAGCAGCAGCAAGCTTGTAGTCGGAGCAG-----GTAGGA	5140
Qy	218	AlaAsnValSerGluAsp-----AsnThrSerValAspValGlyAlaLeuProGlu	234
Db	5141	AACACAGATCTAAGAAAAAATAAACAAACCATCGATGATATAGCAGCTGGAAGAG	5200
Qy	235	ValProGlnIleThrTrpHisIleGluValAsnGlyAla-----	247
Db	5201	AAACAAATGCGATCAGCGTGTGCAACAGTTCGGCTCATCTGAAATTCGCGTGTGTT	5260
Qy	248	---AspGlnProProSerThrProLysLeuSerGluValValLeuLysArgAsnGluAsp	266
Db	5261	GAGGATACACCACTTAAGGCAACCAAGAAATCATAAAGGCATTCGCGTGTGATGGAATGC	5320
Qy	267	GluAsnGlyLysThrGluGluThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro	286
Db	5321	GATTAACGGGTTCATCAGAAAGTATAAATCTGCT-----	5353
Qy	287	AsnProMetSerGlyLysGluArg-----AspGlnValAlaGluGlnCysAsnLeu	303
Db	5354	-----ATGAGTGGGTTCAGCGTAGGAAATCTCGCAAGGTTCGTCTACTCAGTGAGTTG	5407
Qy	304	ThrLysAspProLysProValSerGlyGlnLysCysGluGlnIleCysAsnGluProCys	323
Db	5408	CTTGTTATACAAAACCAAGTGGTGTAGTACATCAGAAA-----	5449
Qy	324	GluGluValValLeuLysArgSerSerLysSerLysArgLysThrAspLysLeuMet	343
Db	5450	GAAGAGTCTGCTTTGAAGAGGAATCAGTTAGAGCTCGAAA-----AGAAAGTTGTTA	5503
Qy	344	LysLysGlnGlnHis---SerLysLysArgThrAlaGlnAlaAspValSerAsp-----	360
Db	5504	CCTGAAACAAATATATGTCAGCGGATATTGAGTACAATGGGTGCAACCTCTGAAAATGCT	5563
Qy	361	AlaLysLeuCys-----	364
Db	5564	TCCAAAAGTCTGACTCTGATCAAGGTAATAGTGAATCAACTAGTGGTGTGACAGA	5623
Qy	365	-----ArgArgLysProLysLysValArgLeuSerGluIleAsn	379
Db	5624	ACTCCATTAAAGGGTAAGCAGAGAAACAGAGANVTTCAGGTGTGTGACAGAGTTGTACCA	5683
Qy	380	AlaAsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAspPro	399

Db	5684	TCATCTCTTGTGAAACTTCACAA-----GAAGGTATCAAGGAGCATGATGATGATCCT	5737
Qy	400	CysGluAspAspArgSerThr-----IlePro	408
Db	5738	AGTAAG-----AGATCAACTCTCTGCGCATCTTTATTCACTGAAACGATCTCTGTTCT	5791
Qy	409	ValProMetGluValSerMetAspIleProValSerAsnHisThrValGlyGluAspGly	428
Db	5792	TGTCCT-----CCGGTACTCAGAGAACA-----GAGAGGAAG	5824
Qy	429	LeuLysSerSerLysAsnLysThrLysArgLysTyrSerAspValValAspAspGlySer	448
Db	5825	CTCAGTTTACCCCAAGAGACAGAGAGAG-----CCTGTAATGATATATGGAAG	5875
Qy	449	SerLeuMet-----AsnTrpLeuAsnGlyLysLys-----LysArgThrGly	462
Db	5876	AGCACTGTGATCAGTTTATAGTAACGCGCATGTGAGGAAGTCAAGTTAATCGCATCTGTT	5935
Qy	463	SerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysValThrPro	482
Db	5936	CCTTCATGACACAGATATCCCAACTCGAGACTTATGATGGGAAAGGGTGGCGGT	5995
Qy	483	ThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAsp-----	499
Db	5996	TTATTTGACAAACCGTTTGGCTTCAGATGGATATTTTCAGAAAATATCTCTCAGGTTAAT	6055
Qy	500	-----ThrAsnMetHisLysThrAspValCysGlnHisValSerGluIleSer	515
Db	6056	GATAAGCCGATAACATCTTTCATTTTTCGAGCAATGATATGAGGTCAAGAGCGCG	6115
Qy	516	ThrGlnArgCys-----SerSerLysGlyLysThrAlaGly---Leu	528
Db	6116	GAACCAACTGCTCTCGAGATTTTGTCTCTCTTCTTAAATCCAGCTCAGGTGATGTTG	6175
Qy	529	SerLysGly-----LysThr	533
Db	6176	AGAAGTGGAGTAGATATTGTGATTCAGAAACAAACCAATACAAACAGATGCTCT	6235
Qy	534	HisSerAlaAlaSerThrLysThrGlyGlyGluSerThrArgAsnGlyGlnAsnIleHis	553
Db	6236	TTCTCGAAGTCTGAGCTAGATACCCCTCTCTCTACTAGTTCGGATTTATCTCG	6295
Qy	554	ValLeuSerAlaGluAspGln---CysGlnMetGluThrGluAsnSerValLeuSerHis	573
Db	6296	GTCTGCAAAAGGTACAAAGCTTTGTCATCTATTTTCTCAATCGCCATGCTTTCTCATTC	6355
Qy	573	-----	573
Db	6356	TGTTTCCAACTTGCACCTATTCTTGTATATTCCAAACCTGCTGCTCTCAGTATGGAT	6415
Qy	574	-----Ala---LysValSerProAlaGluHisAspIleG	584
Db	6416	CAGGATGATAACCGGATTTGATTTTACTTGGCAGGATGCTTCTGTCAGATAGAAAGGGA	6475
Qy	584	IleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysLysGlnLysLeuG	604
Db	6476	AGACTTTATGGTCCCAAGAACATCATGGAGCCACAGAAAGCCAAAGTCCAGTAGAAAGG	6535
Qy	604	IuValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeu	624
Db	6536	AGACTACGACTGAGAGCAAAACAAAC---GATGATATTCATGGAGATAGTGGAGCTCA	6592
Qy	624	euAlaLysAsnGlnHisGluArgGlnLeuMet---ThrGluThrAspCysSerAspIleA	643
Db	6593	TGGCCAAAACCAAGTACGAGAGGTGTCTCCGACAAAGAAAGATGTTAGCAACAAAC	6652
Qy	643	snArgIleGlnSerLysThrThrAlaAspAspCysValIleValAlaAlaLysAspG	663
Db	6653	AGGCATCAAGAAACAGACACACAAATCCAGATGCTCTACTGATGTGATCTCAATGAAA	6712
Qy	663	IySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGln-----LysSerLeuA	681
Db	6713	CCTACGATTAACGGGATCTCACTTGGAGCAACAAACACATCAAGACCAACCAACCGTGT	6772



CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)  
 XX  
 SQ Sequence 5286 BP: 1786 A; 1178 C; 1073 G; 1249 T; 0 U; 0 Other;  
 XX

Alignment Scores:

Alignment Scores:		
Pred. No.:	2.09e-06	Length: 5286
Score:	223.50	Matches: 210
Percent Similarity:	34.23%	Conservative: 173
Best Local Similarity:	18.77%	Mismatches: 412
Query Match:	4.04%	Indels: 325
DB:	5	Gaps: 52

US-09-721-114-2 (1-1057) X AAS73156 (1-5286)

Qy	83	HisAspGln-----LysLysCysAspGluHisLysAlaSerSerProPheSer	99
			---
Db	2236	CATGATGAAGTGGTTGATGTCAAATGCCATTCACACTCTCTTTAGGATGAAGAGGA	2295
Qy	100	ValAlaLysPheArgArgTTPAspCysSerLysCysLeuAspLysLeuLysThrSerAsp	119
			---
Db	2296	AAAGGAAAAATAAGSCATCAT-----ATATCCTGTATTGAAAGTTAAACAACAACAGAA	2349
			---
Qy	120	AsnGlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSer	139
			---
Db	2350	AGTATATCATGATACCC-----ACCATGATCATCAGGAGC	2382
Qy	140	IleThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysValSer---	158
			---
Db	2383	CTC-----ATTGAAGCAAATCAAAAGCAATTCCAAAGTTTCTGAA	2421
Qy	159	-----ProSerThrGlnSerSer-----GlnGlyLys	167
			---
Db	2422	CTTGACACAATTATTATGCTTCCCAAGAAAAATCAAGCAGTTTCTCATACATGGCAGG	2481
Qy	168	AspAlaAspArgSerThrLeuProLysSerValGlnGluGly-----	181
			---
Db	2482	CAGTCAGGAAGTAAATAATATGCTGCTTCATTGAGGAATGGCCACCTCCCTCCAAATC	2541
Qy	182	-----AsnAspSerLysCysAsnAla	188
			---
Db	2542	AAAAATAATGTGGAAGATGCAATGGGGAACTATATGTTAAACAAATTAGTCCCGAGTTCT	2601
Qy	189	ProSerGlyLysAsnGlyAlaAlaGluAlaAsnThrAspSerProMetLysAspLeuGln	208
			---
Db	2602	CCTCAGTCAGCGAATGAATGTTCCAAAGTCCTTTTCAGACTCAGCC-----CTGGAA	2652
Qy	209	GlyProAlaGlnAsnThrAspValAlaAlaAsnValSerGluAspAsnThrSerValAsp	228
			---
Db	2653	GCACCTGAAGCCACAGAGAGAATGCACAAATGTAAAGACGAGTGATCTACTTCGGTTAGA	2712
Qy	229	ValGlyAlaLeuPro-----GluValProGlnIleThrTrpHis	241
			---
Db	2713	AAAGGACCACTTCATTCCTCATCAACAGGGCTATGTCATGCTCCCTCAGGGAGCCACAT	2772
Qy	242	IleGlu-----ValAsnGlyAlaAspGlnProPro	251
			---
Db	2773	GCCTCAACTGGAGAGAAGGAAGAAAAAGCCATTCGACCTCAGCATGATGTTCTGAG	2832
Qy	252	SerThrProLysLeuSerGluValValLeuLysArgAsnGluAspGluAsnGlyLysThr	271
			---
Db	2833	CTAACACCAAGGGCTGGGAGAGAATCATTAGCCCTGTGGAAAGTGACTCATCT-----	2886
Qy	272	GluGluThrLeuValAlaGluGlnCysAsnLeuThrLys-----	284
			---
Db	2887	-----GTTAGAGATTGTTCTTTAACCAAAAGACACACCAACCAAGGAAAAC	2931
Qy	285	-----AspProAsnProMetSerGlyLysGluArgAspGlnValAlaGlu	299
			---

D	b	2932	TTCCAAGAAATACACTGAGAAAGAGGACTTGTCAGGAAAAAGAACGCTGGGAAAAACAATTACAT	2999
Q	y	300	GlnCysAsnLeuThrLysAspProLysProValSerGlyGlnLysCysGluGlnIleCys	319
D	b	2992	AAAGTTAAG--ACGACTAGTACGTTTCTGTTTCTGTTGATGAAGATAATGTAAATGT	3048
Q	y	320	AsnGluPro-CysGluGluValValLeuLysArgSerSerLysSerLys	335
D	b	3049	CTGGAGGTGCTCAATATATATACACTCTACCGAGGAACCCGACGAAAAAATTTCTGTAAAC	3108
Q	y	335	-----	335
D	b	3109	CTCCTTCAACAGTATACACAAATACTAATTTACTTTATAGAAATCACCTCAAGTGGAGACT	3168
Q	y	336	-----ArgLysThrAspLysLysLeuMet-----LysLysGI	346
D	b	3169	GAACACATTTCTTAACCGCTTTTAAAAAAGCAACACAGAAATTTTCTACAGAGCAGCTCA	3228
Q	y	346	ngInHisSerLysLysArgThrAlaGlnAlaAspValSerAspAlaLysLeuLysCysArgAr	366
D	b	3229	GGAACACCTTCATGTGAAATCTAAAGATG-----CCGTCAACTCTGATCAGACG	3279
Q	y	366	GlySerProLysValArgLeuLeuSerGluIleLeuAsnAlaAsnGlnValGluAspSe	386
D	b	3280	CTCACCACCTGAAATATGACTGCCTTCGATTTATCA-AATAGGGGGGCCCTAGCTCCTAC	3338
Q	y	386	ArgSerAspGluValHisArgGluAsnAlaAlaAspProCysGluAspArgSerTh	406
D	b	3339	ATTACAGGAATGGCTTCTGTTGAGGAGCT-----	3369
Q	y	406	rIleProValProMetGluValSerMetAspIleProValSerAsnHisThrValGlyGI	426
D	b	3370	-GTTCTCTCTCCTGAAGAGGAATCTAAAGCTAGAGAGATTTTTCAGATAATTTAGCTAA	3428
Q	y	426	uAspGlyLeuLysSerSerLysValLysThrLysArgLysTyrSerAspValValAspAs	446
D	b	3429	AACACCTCTAGGTGATTCAGAAAAACAGAGGAAGA-----	3465
Q	y	446	pGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysLysArgThrGlySerValHisHi	466
D	b	3466	-----GGCAAAAAGTTGCAAAAGTGAACCCGTCATAC	3497
Q	y	466	sThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysValThrProThrAlaSerTh	486
D	b	3498	TTCAATGATGCTTCAGAGAAAAATGTATCCGAAAGAAAAA---TCTGAATAATTTGTCACAA	3554
Q	y	486	rGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAspThrAsnMetHisLysThrAs	506
D	b	3555	ATCCATTAAATTCAGTAAACAGTGGTCCCTAGTCTTCCAGCT--CTTTCAGAGATTAA	3611
Q	y	506	pValCysGlnHisValSerThLisThrGlnArgCysSerSerLysGlyLysThr--	525
D	b	3612	TATTGGAAATFTCCAAACTAGAAAGATGTTCTGGGAGTGATACAGGGAGTGATAGAGCCAT	3671
Q	y	526	-----AlaGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysTy	541
D	b	3672	TCCATTACTGGAAAGTGGCAAGTGTCCCAAGAAAGATCACATCCACAGCTTAGTGTA	3731
Q	y	541	rGly-----GlyGluSerThrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAs	559
D	b	3732	TGGCTCCAGTGGATCACAGCCTAGGGAAGGCACAGGG-----GACATTGGAAC	3779
Q	y	559	pGlnCysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLys-----	575
D	b	3780	CAACTGCAAAAAAATGACT---AATAAAACACTTTCTCACTCAGAGAGTCAAGTCTTTGC	3836
Q	y	576	-ValSerProAlaGluHisAspIleGlnIleMetSerAspLeuHisGlu-----	591
D	b	3837	CCTTACTCCAGCAATTCGATAAACTACAGCTTGTGTAGGAGACTCAGTCAGATGAACCAA	3896
Q	y	592	-----GlnSerLeuProLysLys-----LysLysLysGlnLysLeuGluVa	605
D	b	3897	CTTAGAGAGTCTGCAGTCTCAACCAAGAGAAATTTACCTCAAGAAAGTCAAGAGCAATAT	3956

QY 605 lThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAl 625  
Db 3957 GACAGAGAGCAGAG-  
QY 625 aLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgI1 645  
Db 3975 TGAAGATGAATGACGAAGTCAGCTTGGGATCAACCTTCCTCTGTAAGGAACACAA- 4032  
QY 645 eGlnSerLysThrThrAlaAspAspCysValIleValAlaAlaLysAspGlySerAs 665  
Db 4033 -AATAAACCAACTGGATGAC-CTAGTAAGCGGGAATAAGATCTTC 4079  
QY 665 pTyr- -AlaSerSerValPheAspThrAsn- 674  
Db 4080 AGTTAAACACAGATGGCAGCCATGCTAAAGCAGCAGCAAAATTCACGCTAAAGATGT 4139  
QY 675 -SerGlnGlnLysSerLeuAlaSer- -GlnSerThrGlnLysGluLeuGlnG1 691  
Db 4140 AAGCCCGAAGACATGATGACTACTATCTTCCCCCAAGTGAAGCAGATCTGGCTTTGA 4199  
QY 691 yHisLeuAlaLeuThrThrGlnGlu- -SerProHisProGlnAs 705  
Db 4200 CCATTATCTCTGGCAGATGGAGTCAACCCACTGTTCCCTGAGCTACTCCAAATC 4259  
QY 705 nPheGlnSerThrGlnGluGlnGln- -ThrHisLeuArgMetGluG1 720  
Db 4260 TGCAGATCCATTGGCGAAGACAGGTGTAGTGAGAACGGAAGCATGTGAAGAAATCCGA 4319  
QY 720 u- -MetValThrIleAl 725  
Db 4320 GAACCTTCTCCCATCTACTACTACCAACAGAGAACCTTCTACACAGCTCAGCAACCA 4379  
QY 725 aAlaSerSerProLeuPheSerHisHisAspAspGlnTyr- -IleAlaGluAlaPr 743  
Db 4380 GAAGTCTTAACAGCATTTCAACAGCAGCATCAGATGATTTAAACACGCTCTCAGAAATCACC 4439  
QY 743 oThrGluHisTrpGlyArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaTh 763  
Db 4440 ATCAACGATGAGATTTCTAAGAC- -GTCAC 4469  
QY 763 fThrArgAsnSerProAlaAlaThrCysGlyAlaGlnPheArgProGlyIleGlnAlaVa 783  
Db 4470 AGCAGCTCAGAAATTTAGTAAGAGAAATCAGGAGCCCA- -TCACCCAT 4514  
QY 783 lAspLeuThrSerThrHisValMetGlySerSerAsnTyrAlaSerArgGlnProVa 803  
Db 4515 CACATTCACAGCTCAGGAGCAGCAATCTCTGACAACTCAGAGAGGCTGAGCCCTCC 4574  
QY 803 lIleAlaProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPh 823  
Db 4575 TTTT- -CCACTGGAGCT- -GCACAGAAA- -4599  
QY 823 eProSerThrIleAlaThrMetGluAlaSerLysLeuCysAspArgAsnAlaGlyG1 843  
Db 4600 -TCTAGATAGCAGCTCAGCTGGCAGTTTCTGACGAAACAAAGAGTGTCTTCATC 4655  
QY 843 nValValLeuTyrProLysGluSerMetProAlaThrHisLeuLeuArgMetMetAspPr 863  
Db 4656 TCTG- -GAGTGGGAACCTGAGCCACACTCTATCGT- -TC 4691  
QY 863 oSerThrLeuAlaSerPheProAsnTyrGlyThrSerSerArgAsnGlnMetGluSerG1 883  
Db 4692 AAAGAGTTTAAAGAGCATTAATGTTTCATGGCGATCTACTACGAAAGAGC- -4740  
QY 883 nLeuHisAsnSerGlnTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSertyrG1 903  
Db 4741 - -CATCTCCAAAGTCAGGAGGCCATTTTCTGAAAGCATCTCTATT- -4788  
QY 903 ySerAsnLeuAsnGlyLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHi 923  
Db 4789 -CACAATGCCCTGAGTCGACTGACCTTGGGAATGAATTTCTCTCAACAATGGGTACAG 4847

QY 923 sAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlySerLeuLeuG1 943  
Db 4848 TCGAAGATTCAGATCTTTTCTGAACTCCCTCTCTGTGAT- -GGAAAT- -4893  
QY 943 nLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyVa 963  
Db 4894 - -GAAGTGGCTTATCGCAGCGGACAAAACAGGTCCAGGCTGTGCAAT 4943  
QY 963 lSerThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSe 983  
Db 4944 ATCTATA- -TACAGACCTATCGACTA 4967  
QY 983 rGlyMetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerAlaAs 1003  
Db 4968 TGGGATCTTTGGGAAGAACCAACAGTTAGCTTCTTGAATAATGTAAAGAGGTCA- -5022  
QY 1003 pPheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysMetVa 1023  
Db 5023 - -CTTACACAGCAAGATATTGCAACCAAGTTTCTTAAGAACCCCTGGC- -5070  
QY 1023 lHisProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAsnPro 1041  
Db 5071 - -TTCTTAAAGATGATTTG- -AGAACCCCT 5097

RESULT 11  
ABL68623  
ID ABL68623 standard; DNA; 10452 BP.  
XX ABL68623;  
XX 15-MAY-2002 (first entry)  
DT XX  
DE XX  
DE XX  
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX Homo sapiens.  
XX WO200194623-A2.  
XX 13-DEC-2001.  
XX 30-MAY-2001; 2001WO-US010838.  
XX 05-JUN-2000; 2000US-0209473P.  
XX 05-JUN-2000; 2000US-0209531P.  
XX 18-SEP-2000; 2000US-0233133P.  
XX 18-SEP-2000; 2000US-0233617P.  
XX 20-SEP-2000; 2000US-0234009P.  
XX 20-SEP-2000; 2000US-0234034P.  
XX 20-SEP-2000; 2000US-0234052P.  
XX 22-SEP-2000; 2000US-0234509P.  
XX 22-SEP-2000; 2000US-0234567P.  
XX 25-SEP-2000; 2000US-0234923P.  
XX 25-SEP-2000; 2000US-0234924P.  
XX 25-SEP-2000; 2000US-0235077P.  
XX 25-SEP-2000; 2000US-0235082P.  
XX 25-SEP-2000; 2000US-0235134P.  
XX 25-SEP-2000; 2000US-0235280P.  
XX 26-SEP-2000; 2000US-0235637P.  
XX 26-SEP-2000; 2000US-0235638P.  
XX 27-SEP-2000; 2000US-0235711P.  
XX 27-SEP-2000; 2000US-0235720P.  
XX 27-SEP-2000; 2000US-0235840P.  
XX 27-SEP-2000; 2000US-0235863P.  
XX 28-SEP-2000; 2000US-0236028P.  
XX 28-SEP-2000; 2000US-0236032P.  
XX 28-SEP-2000; 2000US-0236033P.  
XX 28-SEP-2000; 2000US-0236034P.  
XX 28-SEP-2000; 2000US-0236109P.

PR	28-SEP-2000;	2000US-0236111P.
PR	29-SEP-2000;	2000US-0236843P.
PR	29-SEP-2000;	2000US-0236891P.
PR	02-OCT-2000;	2000US-0237172P.
PR	02-OCT-2000;	2000US-0237173P.
PR	02-OCT-2000;	2000US-0237278P.
PR	02-OCT-2000;	2000US-0237294P.
PR	02-OCT-2000;	2000US-0237295P.
PR	02-OCT-2000;	2000US-0237316P.
PR	03-OCT-2000;	2000US-0237425P.
PR	03-OCT-2000;	2000US-0237598P.
PR	03-OCT-2000;	2000US-0237604P.
PR	03-OCT-2000;	2000US-0237605P.
PR	03-OCT-2000;	2000US-0237606P.
PR	01-NOV-2000;	2000US-0244867P.
XX	01-NOV-2000;	2000US-0245084P.
XX	(AVAL-) AVALON PHARM.	
XX		
XX	Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;	
PI	Soppet DR, Weaver Z;	
XX	WPI; 2002-188264/24.	
XX		
PT	Screening for anti-neoplastic agent involves exposing cells to a chemical	
PT	agent to be tested for anti-neoplastic activity, and determining a change	
PT	in expression of a gene of a signature gene set.	
XX		
PS	Claim 1; SEQ ID NO 6960; 44pp; English.	
XX		
CC	The present invention describes a method (M1) for screening for an anti-	
CC	neoplastic agent. The method involves exposing cells to a chemical agent	
CC	to be tested for anti-neoplastic activity, determining a change in	
CC	expression of at least one gene (I) of a signature gene set, where (I)	
CC	comprises a sequence (S) selected from 8447 sequences (given in AB161664	
CC	to AB170110), or is at least 95% identical to (S), where a change in	
CC	expression is indicative of anti-neoplastic activity. (I) has cytostatic	
CC	activity and can be used in gene therapy. M1 can be used for screening an	
CC	anti-neoplastic agent, and can be used for producing a product which is	
CC	the data collected with respect to the anti-neoplastic agent as a result	
CC	of M1, and the data is sufficient to convey the chemical structure and/or	
CC	properties of the agent. M1 can be used in the treatment of cancer such	
CC	as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,	
CC	prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell	
CC	cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous	
CC	cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms	
CC	tumour	
XX		
SQ	Sequence 10452 BP; 3702 A; 1658 C; 2212 G; 2880 T; 0 U; 0 Other;	
	Alignment Scores:	
Pred. No.:	5,61e-06	Length: 10452
Score:	223.00	Matches: 172
Percent Similarity:	34.31%	Conservative: 108
Best Local Similarity:	21.08%	Mismatches: 312
Query Match:	4.04%	Indels: 225
DB:	6	Gaps: 33
US-09-721-114-2 (1-1057) x ABL68623 {1-10452}		
Qy	62 AtgGlyTYrValAlaLeuGlnLysLyAspProLysPheCysSerLysSerArgIle 81	
Dd	: : : : :           : : : : :           : : : : :           : : : : :	
	1547 AGGCCTCACTTCGATTGGAGAGACACTTAATCCGAGTTTCGCGGTGATGCTGTA 1606	
Qy	82 PheHisAspGlnLysLyCysAspGluHISylvaIsaSerSerProPheSerVal--- 100	
Dd	: : : : :           : : : : :           : : : : :           : : : : :	
	1607 ---AACAAAGAGAAAATAACCAAAGAGCATAAAGTCAATAGATCTTAAGTTTGAAACAAAA 1663	
Qy	101 -----AlaIysPheArgTrpAspCysSerLysCysLeuasp 113	
Dd	1664 GCACGAAGAGAAAAACCTTGTGCTTTTGGAAAGAAGGATATTTCCAAGTCAGAGCT 1723	
Qy	114 LysLeu----- 115	



CC disease, ulcerative colitis, periodontal disease; also bacterial  
CC infection, viral infection, parasitic infection, protozoal infection,  
CC fungal infection and M5 is useful for treating one of the above  
CC conditions. The present sequence represents a gene differentially  
CC expressed in granulocytes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 10452 BP; 3702 A; 1658 C; 2212 G; 2880 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5.61e-06 Length: 10452  
Score: 223.00 Matches: 172  
Percent Similarity: 34.31% Conservative: 108  
Best Local Similarity: 21.08% Mismatches: 312  
Query Match: 4.04% Indels: 225  
DB: 6 Gaps: 33

US-09-721-114-2 (1-1057) x ABK84422 (1-10452)

QY 62 ArgGlyTyrValAlaLeuLeuGlnLysLysAspProLysPheCysSerLeuSerArgIle 81  
DB 1547 AAGGCTCATCTTCGATTTGGGAAGAGACTTAAATTCGAGTTTCGAGCGATGGATGCTGTA 1606  
QY 82 PheHisAspGlnLysLysCysAspGluHisLysAlaSerSerProPheSerVal--- 100  
DB 1607 ---AACAAAGAGAAAATACCAAGAGACATAAAGTCTATAGATGCTAAAGTTTGAACAAAA 1663  
QY 101 -----AlaLysPheArgArgTTrpAspCysSerLysCysLeuAsp 113  
DB 1664 GCACGAAAGAGAAAACCTTGTGCTTTGGAAAAGAGATATTCAAGTCAGAGCT 1723  
QY 114 LysLeu----- 115  
DB 1724 AAACCTTCAGAAAAACAGGTAGATAGTGCACATGCATCAGAAATGTTCCACAGAGGAA 1783  
QY 116 -----LysThrSerAspAsnGlyThrAlaPro 124  
DB 1784 CAAGAACAAATAAAGTACCGGTGGTGAACATAAGAAATCTGTAGAAAAGAACCT 1843  
QY 125 ArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIleThrPheValArg 144  
DB 1844 CAATATGAACCTGCC-----AACACTTCTGAAGATTTAGACATGGATTTGTGTCT 1894  
QY 145 SerThrPheValProAlaSerValGlySerGln-----LysValSerPro 159  
DB 1895 -----GTTCTCTTCCAGTTCAGAGACATTTTGAAGATCTTGAGACTGCTATG 1945  
QY 160 SerThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGln 179  
DB 1946 GAAGTTCAAGATTCAGTTGATCATCAGGGGATGGCAGCAGTGAAGTGAACAGAGATG 2005  
QY 180 GluGlyAsnAspSerLysCysAsnAlaProSerGlyLysAsn-----GlyAlaAlaGlu 197  
DB 2006 GAGAGTTTCATCTGTAAAAATTAATATTCTTCAAAAGACAAACAGAGAGGTATTAAATCA 2065  
QY 198 AlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAla 217  
DB 2066 AAAACTACACTAAAGTAACAAAGAAATTA-----TATGTAAACTC 2107  
QY 218 AlaAsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGln 237  
DB 2108 ACTCCTGTCTCCCTTCTAATTCCTCAATTCCTCAATTCCTCAATTCCTCAATTCCTCA 2167  
QY 238 IleThrTrpHisIleGluValAsnGlyAlaAspGlnProSerThrProLysLeuSer 257  
DB 2168 -----GATAAAGATGGCTATATAAAGTTGTGGTCTGAACCCCAAGTTAGAG 2212  
QY 258 GluValValLeuLysArgAsnGluAspGluAsnGly-----LysThrGluGluThr 274  
DB 2213 AAATGTGGACTTCGACAGGAAAACAGTGATATAGCAATTTGGTTGGAATGAGTTTCA 2272

QY 275 LeuValAlaGluGlnCysAsnLeuThrLysAspProAsnProMetSerGlyLysGluArg 294  
DB 2273 TTACTTTTAGAGGAATCTGATCTTCGAGATCCCAAGGTTAAAGACTACACCTTTGAGG 2332  
QY 295 AspGlnValAlaGluGlnCysAsnLeuThrLysAspProLysProLysGlyGlnLys 314  
DB 2333 CGACCGACAC-----GAACTAAACCTCTGTAACATCTAATTC 2368  
QY 315 CysGluGlnIleCysAsnGluProCysGluGluValValLeuLysArgSerLysSer 334  
DB 2369 GATGAGAA---TGTAATGAACAGTTAAGAG-----AAACAAACTATCAGTTCCA 2419  
QY 335 LysArgLysThrAspLysLysLeuMetLysLysGlnGlnHisSerLysLysArgThrAla 354  
DB 2420 GTGAGAAAAAAGATAAGCGT----- 2440  
QY 355 GlnAlaAspValSerAspAlaLysLysCysArgLysProLysLysVal-----Arg 372  
DB 2441 -----AATCTCTGACAGTCTATAGTAAATCTTAAGCTTAATTAATTCGCAAAATCT 2494  
QY 373 LeuLeuSerGluIleIleAsnAlaAsnGlnValGluAspSerArg-----SerAsp 389  
DB 2495 AAGCAATCAGACACTGTGTGATCAAAATTCAGATCTCTGATGAATGTAGCAATCTCTCAA 3554  
QY 390 GluValHisArgGluAsnAlaAlaAspProCysGluAspAspArgSerThrIleProVal 409  
DB 2555 GAGGTGAGCAGGATGAGTCACAGTCTCTTCTTCAGATCTGATATTAATGAATTCATACA 2614  
QY 410 ProMetGluValSerMetAspIleProValSerAsnHisThrValGlyGluAspGlyLeu 429  
DB 2615 ACCATAAGACTTGTATGATTTAAAGACTCAG-----GCGGGGAAGAT----- 2659  
QY 430 LysSerSerLysAsnLysThrLysArgLysTyrSerAspValValAspAsp----- 446  
DB 2660 -----GATAAAGGAAAAAGGAAACGAAAGTCTTACATCTGGCTCAGATTTTGATACT 2713  
QY 447 -----GlySerSerLeuMetAsnTrpLeuAsnGlyLysLysArg----- 460  
DB 2714 AAAAAGGCAAAATCAGTAAAGCTCTATAATTTCTAAAGAAACGACAAACCCAGTCT 2773  
QY 461 -----ThrGlySerValHisHisThrValAlaHisPro 471  
DB 2774 GAGTCTTCTAATATATCACTCAGAAATTAGAAAAGAGATAAGAGATCAGTAAATTTGT 2833  
QY 472 AlaGlyAsnLeuSerAsnLysLysValThrProThrAlaSer---ThrGlnHisAspAsp 490  
DB 2834 GTGCGCAGAACCCCAAAAAGGAATTCGAATCAAAAGATTTTGACTCTTCTTGAGAT 2893  
QY 491 GluAsnAspThrGluAsnGlyLeuAspThrAsnMetHisLysThr---AspValCysGln 509  
DB 2894 GAGAAACACAGCAAAAAGGAATGATTAATCAAGGCGCAAAAATTTGAAGACCTCACA 2953  
QY 510 HisValSerGluIleSerThrGlnArgCysSerSerLysLysThrAlaGlyLeuSer 529  
DB 2954 GAAGGATTCATCTGATGATCTGAAAAGA---AAAACAGAGAGAGAGACTTCTCTTCAGCA 3010  
QY 530 LysGly-----Lys 532  
DB 3011 GAAGGCACAGTTGATAAAGACACAGCACCATCATGGAATTAAGAGATCGACTTCTTAAGA 3070  
QY 533 ThrHisSerAlaAspSerThrLysTyrGlyGlyGluSerThrArgAsnGlyGlnAsnIle 552  
DB 3071 CAGCAAGCAAGTGTCTCCACT-----GATGTGTGCGATAAGCTTCTTGGGAAAGAGCA 3124  
QY 553 HisValLeuSerAlaGluAspGlnCysGlnMetGluThrGluAsnSerValLeuSerHis 572  
DB 3125 AGTTTACTTCTTCGAGTAGTAAAAAGTGTGTAACCTAAGAAAAG-----AGCAAG 3178  
QY 573 SerAlaLysValSerProAlaGluHisAspIleGlnIleMetSerAspLeuHisGluGln 592  
DB 3179 CATCTCAAAACCAACATGTAATAAAGTACAGGATGGCTTATCTGATATTCAGAGAAA 3238  
QY 593 SerLeuProLysLysLys-----LysLysGlnLysLeuGlu 604



Db 3239 TTCCTAAAGAACACGAGCGATGAACTTCTGAAGTATGATAAAAGCAGAGCAAAAG 3298  
 Qy 605 ValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeu 524  
 Db 3299 GGAACCTGAAGAAAAGAAA---CCTTCGACTTTAAGAAAGTA 3343  
 Qy 625 AlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSerAsp 641  
 Db 3344 ATTAAA-----ATGGAACAACAGTATGATCTTCATCTGATGGCACTGAAGAAGTACT 3397  
 Qy 642 -----IleAsnArgIleGlnSerLysThrThr 550  
 Db 3398 GAGCGAGAGAATAATTTGTCAATTTCTTAAGGCGCATAAACAAATTAAGAAATGGAACAAC 3457  
 Qy 651 AlaAspAspCysValIleValAlaLysAspGlySerAspThrAlaSerVal 670  
 Db 3458 -----GATGGAGAAAGAAAAGTAATAATA 3484  
 Qy 671 PheAspThrAsnSerGlnGlnLysSerLeuAlaSerGlnSerThrGlnLysGluLeuGln 690  
 Db 3485 AGAGATATAAATCTCTAAAGAAAGAGATGAATATCTGATGCTGAGAAGTCAA--CAG 3542  
 Qy 691 GlyHisLeuAlaLeuThrThrGlnGluSerProHisProGlnAsnPheGlnSerThrGln 710  
 Db 3543 GGA----- 3545  
 Qy 711 GluGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerSerProLeu 730  
 Db 3546 -----AAGGAGATAGTTGTGACTCT 3566  
 Qy 731 PheSerHisAspAspGlnThrIleAlaGluAlaProThrGluHis-TripGlyArgly 750  
 Db 3567 -----CAGAGATAAAAGAGTAAGATGAGCATATGTTAGAGAGA 3608  
 Qy 750 sAspAlaLysLysLeuThr-TripGluGlnPheLysAlaThrThrArg 765  
 Db 3609 AGAAAGGTGCAAGTTGCTTGGAAAGAGTTCAGAGGAAGAGACAGA 3654

RESULT 13  
 ADE53841  
 ID ADE53841 standard; cDNA; 11167 BP.  
 AC ADE53841;  
 DT 29-JAN-2004 (first entry)  
 DE Human prostate cancer cDNA #188.  
 KW Human; prostate cancer; ss; cDNA combination; differential expression;  
 KW gene.  
 OS Homo sapiens.  
 XX US2003190640-A1.  
 PN 09-OCT-2003.  
 PD 29-MAY-2002; 2002US-00252157.  
 PF 31-MAY-2001; 2001US-0295048P.  
 PR (FARI/) PARIS M.  
 XX (PEAR/) PEARSON C I.  
 PA Paris M, Pearson CI;  
 PI WPI; 2003-831618/77.  
 DR New combination comprising cDNAs that are differentially expressed in  
 XX prostate cancer, useful for diagnosing, treating or monitoring the  
 PT progression of treatment of prostate cancer.

PS Claim 1; SEQ ID NO 188; 42pp; English.

XX The invention relates to a combination comprising a number of cDNAs  
 CC expressed in prostate cancer. The invention also relates to a method for  
 CC detecting differential expression of one or more cDNAs in a sample  
 CC containing nucleic acids by hybridising a substrate with the nucleic  
 CC acids, thus forming one or more hybridisation complexes, detecting  
 CC hybridisation complex formation and comparing the complexes formed with  
 CC standard complexes, where differences between the standard and the sample  
 CC complex formation indicate differential expression of cDNAs in the  
 CC sample. The differential expression is diagnostic of prostate cancer. The  
 CC invention also relates to proteins and antibodies related to the cDNAs.  
 CC The combination is useful for diagnosing, treating or monitoring the  
 CC progression of treatment of prostate cancer. The antibodies are useful  
 CC for detecting prostate cancer. This sequence represents a human prostate  
 CC cancer cDNA of the invention.

SQ Sequence 11167 BP; 3907 A; 1794 C; 2322 G; 3141 T; 0 U; 3 Other;

Alignment Scores:  
 Pred. No.: 6.12e-06 Length: 11167  
 Score: 223.00 Matches: 172  
 Percent Similarity: 34.31% Conservative: 108  
 Best Local Similarity: 21.08% Mismatches: 312  
 Query Match: 4.04% Indels: 225  
 DB: 9 Gaps: 33

US-09-721-114-2 (1-1057) x ADE53841 (1-11167)

Qy 62 ArgGlyTyrValAlaLeuLeuGlnLysLysAspProLysPheCysSerLeuSerArgIle 81  
 Db 1425 AAGGCTCATCTTCGATTGGAGAAGACCTTAATTCGAGTTTCGAGCGATGCTGTA 1484  
 Qy 82 PheHisAspGlnLysLysCysAspGluHisLysAlaSerSerSerProPheSerVal--- 100  
 Db 1485 ---AACAAAGAGAAAATACCAAGAGCATAAAGTCATAGATGCTAGTTTGAACAAAA 1541  
 Qy 101 -----AlaLysPheArgTrpAspCysSerCysLysCysLeuAsp 113  
 Db 1542 GCACGAAAGGAGAAAACCTTGCTTGTGAAAGAGGATATTTCAAGTCAAGACCT 1601  
 Qy 114 LysLeu----- 115  
 Db 1602 AAACCTTCAAGAAAACAGGTAGTAGTGACACATGCATCAGATGTTCCACAGAGAA 1661  
 Qy 116 -----LysThrSerAspAsnGlyThrAlaPro 124  
 Db 1662 CAAGAACAAATATAAGTACCGGTGGTGAACATAGAATCTGATAGAAAGAGACCT 1721  
 Qy 125 ArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIleThrPheValArg 144  
 Db 1722 CAATATGAACCTGCC-----AACACTTCTGAAGATTTAGACATGATGTTGCT 1772  
 Qy 145 SerThrPheValProAlaSerValGlySerGln-----LysValSerPro 159  
 Db 1773 -----GTTCTTCTCTCAGTCCAGAGACATTTTGGAAATCTTGAGACTGCTATG 1823  
 Qy 160 SerThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGln 179  
 Db 1824 GAAGTTCAGATTGATTCATCAGGGGATGCGACAGTGGAACTGAACAGAAAGTG 1883  
 Qy 180 GluGlyAsnAspSerLysCysAsnAlaProSerGlyLysAsn-----GlyAlaAlaGlu 197  
 Db 1884 GAGAGTTCATCTGTAAATTTAAATATTTCTTCAAAGACACACAGAGAGGTATTAAATCA 1943  
 Qy 198 AlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAla 217  
 Db 1944 AAAACTACAGCTAAAGTAACAAAGAAATTA-----TATGTTAAACTC 1985  
 Qy 218 AlaAsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGln 237  
 Db 1986 ACTCCTGTTCCCTTTCTAAATTCCTCCCAATTAAAGTGTGCTGATGTCAGAGAGTTCCACAA 2045





QY 120 AsnGlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSer 139  
 Db : : : : :  
 QY 2955 AGTATATCAGTACC-----ACAGTGATCAGCAGC 2987  
 Db : : : : :  
 QY 140 IleThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysValSer--- 158  
 Db : : : : :  
 QY 2988 CTC-----ATTGAAGCAAAATCAAGCAAAATTCGAAGTTCTCGAA 3026  
 Db : : : : :  
 QY 159 -----ProSerThrGlnSerSer-----GlnGlyLys 167  
 Db : : : : :  
 QY 3027 CTTGACACATTTATGCTACCTTGGCCAGAAATCAAGCAGTTTCTCATCATGCGCAG 3086  
 Db : : : : :  
 QY 168 AsnAlaAspArgSerThrLeuProLysSerValGlnGluGly----- 181  
 Db : : : : :  
 QY 3087 CAGTCAGGAAGTAAATAATGCTGCTTCATTGAGGAATGGCCACCTCCCTCCCAATC 3146  
 Db : : : : :  
 QY 182 -----AsnAspSerLysCysAsnAla 188  
 Db : : : : :  
 QY 3147 AAAAATAATGTGGAGATGCAATGGGMACTATATGTAAACAATTTAGTCCCACTTCT 3206  
 Db : : : : :  
 QY 189 ProSerGlyLysAsnGlyAlaAlaGluAlaAsnThrAspSerProMetLysAspLeuGln 208  
 Db : : : : :  
 QY 3207 CCGTGAATCAGCAATGAATGTCACAAAGTCCCTTTCAGACTCAGCC-----CTGGAA 3257  
 Db : : : : :  
 QY 209 GlyProAlaGlnAsnTyrAspValAlaAlaAsnValSerGluAspAsnThrSerValAsp 228  
 Db : : : : :  
 QY 3258 GCACCTGAGCCACAGAGAGATGACAAATGTAAAGCAGTGTGATCTCTCCGTAGA 3317  
 Db : : : : :  
 QY 229 ValGlyAlaLeuPro-----GluValProGlnIleThrTrpHis 241  
 Db : : : : :  
 QY 3318 AAAGGACCCTTCCATTCCCTCATCAACAGGGCTATGTGTCCTCCAGGGAGCCACAT 3377  
 Db : : : : :  
 QY 242 IleGlu-----ValAsnGlyAlaAspGlnProPro 251  
 Db : : : : :  
 QY 3378 GCCTCAACTGGAGAGAGAGAGAAAGCCATTTGACCTCAGCGCATGGATCTTCGAG 3437  
 Db : : : : :  
 QY 252 SerThrProLysLeuSerGluValAlaGluGlnCysAsnLeuThrLys 271  
 Db : : : : :  
 QY 3438 CTAACACCAAGGGCTGGGAGAGATCATATGACCTGTGGAAAGTGACTCATCT----- 3491  
 Db : : : : :  
 QY 272 GluGluThrLeuValAlaGluGlnCysAsnLeuThrLys----- 284  
 Db : : : : :  
 QY 3492 -----GTTAGAGATTGTTCTTTAAACCAAAAGACACACACAAAGGAAAC 3536  
 Db : : : : :  
 QY 284 ----- 284  
 Db : : : : :  
 QY 3537 TTCAGAATACACTGGAAGAGGGTAAATGGCTGCTCCAGAGAGAGTATTGCT 3596  
 Db : : : : :  
 QY 285 -----AspProAsnPro-----MetSerGlyLysGluArgAspGln 296  
 Db : : : : :  
 QY 3597 CTTTCAAAATGAAGACCCCTTTACCTTTTGTCTCAGACTTGTTCAGAAAGAACCTGGGAA 3656  
 Db : : : : :  
 QY 297 ValAlaGluGlnCysAsnLeuThrLysAspProLysProValSerGlyGlnLysCysGlu 316  
 Db : : : : :  
 QY 3657 ACATTACATAAAGTTAAG---AOCAGTAGTACGTTTCTGTTCTGTGTGATGATGATAAT 3713  
 Db : : : : :  
 QY 317 GlnIleCysAsnGluPro-CysGluGluValValLeuLysArgSerSerLysSerLys-- 335  
 Db : : : : :  
 QY 3714 GTAAATGTCTGGAGGTGGTCTCAATATATTACACTCTACCGAGAAACCCAGCAAAA 3773  
 Db : : : : :  
 QY 335 ----- 335  
 Db : : : : :  
 QY 3774 TTCTGTAACCTCTTCAACAGTATACAAAATACTAATTTACTATPAGAAATCACCTCAA 3833  
 Db : : : : :  
 QY 336 -----ArgLysThrAspLysLysLeuMet----- 343  
 Db : : : : :  
 QY 3834 GTGAGACTGAAACATTTCTTAAGCTTTTAGAAAAGCAACAGATTTATTCTACACGA 3893  
 Db : : : : :  
 QY 344 -LysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAspAlaLysLe 363  
 Db : : : : :  
 QY 3894 GAGCAGTCAGGAACACCTTCTATGTGAAAATCTAAAGATGT-----CCGTCAACTCT 3944

QY 363 uCysArgArgLysProLysLysValArgLeuLeuSerGluIleLeuAsnAlaAsnGlnVa 383  
 Db : : : : :  
 QY 3945 GATCAGACGCTCACCACCTCAAAATATGCTGCTCCGATTATCA-AATAGGGGGCCCT 4003  
 Db : : : : :  
 QY 383 lGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspProCysGluAspAs 403  
 Db : : : : :  
 QY 4004 AGCTCTTACATTTACAGAAATGGCTTCTGTGTGAGCAGCT----- 4043  
 Db : : : : :  
 QY 403 pArgSerThrIleProValProMetGluValSerMetAspIleProValSerAsnHisTh 423  
 Db : : : : :  
 QY 4044 -----GTTTCTCTCTCTGAAGAGAAATCTAAAGCTAGAGAGATTTTTCAGATAA 4093  
 Db : : : : :  
 QY 423 rValGlyGluAspGlyLysSerSerLysAsnLysThrLysArgLysTyrSerAspVa 443  
 Db : : : : :  
 QY 4094 TTAGCTTAAACACCTCTAGTGTATTCAGAAACAAAGAAAGAAAGA----- 4139  
 Db : : : : :  
 QY 443 lValAspAspGlySerSerSerLeuMetAsnTrpLeuAsnGlyLysLysLysArgThrGlySe 463  
 Db : : : : :  
 QY 4140 -----GGCAAAAAGTTGCAAAAGTGAAC 4162  
 Db : : : : :  
 QY 463 rValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysValThrProTh 483  
 Db : : : : :  
 QY 4163 CCGTGCATCTTCAATGATGCTTCAGAGAAAATATGATCCGAAGAAAAA---TCTGAAAA 4219  
 Db : : : : :  
 QY 483 rAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAspThrAsnMethi 503  
 Db : : : : :  
 QY 4220 TTGTCAACAACTCATTAATTCAGTAAACAGTGGTCCCTCTAGCTTCTCCAGCT---CTTTC 4276  
 Db : : : : :  
 QY 503 sLysThrAspValCysGlnHisValSerGluIleSerThrGlnArgCysSerSerLysGln 523  
 Db : : : : :  
 QY 4277 AGAAGTTATATATGGAAATTCCAAACTAGAAAGATTTCTGGAGTGTACAGGGAGTGG 4336  
 Db : : : : :  
 QY 523 vLysThr-----AlaGlyLeuSerLysGlyLysThrHisSerAlaAlaSe 538  
 Db : : : : :  
 QY 4337 TAGGCCATTCATTTACTGGAAAGTGGCAAGTGTCCCAAGAAAGATCACACATCCACAG 4396  
 Db : : : : :  
 QY 538 rThrLysTyrGly-----GlyGluSerThrArgAsnGlyGlnAsnIleHisValLeuSe 556  
 Db : : : : :  
 QY 4397 TGTAGTGTATGCTCCAGTGGATCACAGCTAGGGAAGGCAGAGG-----GA 4444  
 Db : : : : :  
 QY 556 rAlaGluAspGlnCysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLys-- 575  
 Db : : : : :  
 QY 4445 CATTTGAACCAACTGCGCAAAAATGACT---AATAAAACACTTTCTCCTCAGAGTCA 4501  
 Db : : : : :  
 QY 576 -----ValSerProAlaGluHisAspIleGlnIleMetSerLeuHisGlu-- 591  
 Db : : : : :  
 QY 4502 AGCTTTGGCCCTTACTCCAGCATTCATATAACATACAGCTTGTGTGAGGACTCAGTCAGA 4561  
 Db : : : : :  
 QY 592 -----GlnSerLeuProLysLys-----LysLysLysGlnLys 502  
 Db : : : : :  
 QY 4562 TGAACCAAACTTAGAGAGTCTGCGAGTCTGAACCAAGAGAAATTACCTCAAGAGAGTCAGA 4621  
 Db : : : : :  
 QY 602 sLeuGluValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGln 622  
 Db : : : : :  
 QY 4622 GGCAATATGACAGAGAGAGAGAG----- 4646  
 Db : : : : :  
 QY 622 uLeuLeuAlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSerAspIl 642  
 Db : : : : :  
 QY 4647 -----GCTGAAGATGAATAATGCAGAAAGTCAGCTTGGCATCAACCTTCTCCTCGAAG 4699  
 Db : : : : :  
 QY 642 eAsnArgIleGlnSerLysThrAlaAspAspAspCysValIleValAlaAlaLysAs 662  
 Db : : : : :  
 QY 4700 AAACAAA-----AATAAAACCAACTTGGATGAC-----CTAGTAAGGGGAAAA 4744  
 Db : : : : :  
 QY 662 pGlySerAspTyr-----AlaSerSerValPheAspTh 673  
 Db : : : : :  
 QY 4745 TAGATCTTCTCAGTTTAAACACAGATTTGGCAGCCATGTCTAAAGCCAGCAGAAATTTCCAG 4804  
 Db : : : : :  
 QY 673 rAsn-----SerGlnGlnLysSerLeuAlaSer-----GlnSerThrGlnLysGln 688  
 Db : : : : :  
 QY 4805 TAAAGATGTAAGCCCAAGAGACATGTAGTACTACTTCTCCCAAGTGAAGGAGCATC 4864  
 Db : : : : :  
 QY 688 uLeuGlnGlyHisLeuAlaLeuThrThrGlnGlu-----SerProHi 702

Db 4865 TGGCTTTGACCAATTATCTCTGGCAGTGCAGTGCACCCCACTTCCCTGAGCCTAC 4924  
QY 702 sProGlnAsnGlnSer-----ThrGlnGlnGlnThrHisLeuAr 717  
Db 4925 TCCAAAATCTGCAGAGTCCATTGGCGAAGACAGTGTGAGTGAGAACAGAAAGCATGTGAA 4984  
QY 717 gMetGluGlu-----MetVa 722  
Db 4985 GAATCCGAGAACCTTCTCCCAATTACTGTACTACCCACAGAGAACCTTCTACACAGT 5044  
QY 722 lThrIleAlaAlaSerProLeuPheSerHisHisAspGlnTyr-----IleAl 740  
Db 5045 CAGCAACCCAGAGTCTTACAGCAGCTTTTACACAGCATCAGATGAGTTTAAAAACGTCTC 5104  
QY 740 aGluAlaProThrGlnHisTrpGlyArgLysAspAlaLysLysLeuThrTrpGluGlnPh 760  
Db 5105 AGAATCACCATCAAGCATGAGAAATTTCTTAAGAC----- 5138  
QY 760 eLysAlaThrThrArgAsnSerProAlaAlaThrCysGlyAlaGlnPheArgProGlyI 780  
Db 5139 ----GTCACAGCAGCTCAGAAATTTAGTAAGAGAATCAGGAGCCCA----- 5180  
QY 780 eGlnAlaValAspLeuThrSerThrHisValMetGlySerSerSerAsnTyrAlaSerAr 800  
Db 5181 -TCACCCATCACAATTCACCGAGCTCAGGAGCAGAGAAATCTCTGCAATCAGAGGAGGCT 5239  
QY 800 gGlnProValIleAlaProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAl 820  
Db 5240 GAGCCCTCTCTTT---CCACTGAGCCT---GCACAGAAA----- 5273  
QY 820 aArgAsnPheProSerThrIleAlaThrMetGluAlaSerLysLeuLysCysAspArgAs 840  
Db 5274 -----TCTAGAGTAAGACAGTCCACTGGCCAGTTTCTCGACCAACAAAGGAG 5320  
QY 840 nAlaGlyGlnValValLeuTyrProLysGluSerMetProAlaThrHisLeuLeuArgMe 860  
Db 5321 TGCTTCATCTCTG-----GAGTGGGAACCTGAGCCACACCTCTATCGT--- 5363  
QY 860 tMetAspProSerThrLeuAlaSerPheProAsnTyrGlyThrSerSerArgAsnGlnMe 880  
Db 5364 -----TCAAGAGTTTAAAGACATTAATGTTTCATGGCGATCTACTACGAAAAAGC-- 5414  
QY 880 tGluSerGlnLeuHisAsnSerGlnTyrAlaHisAsnGlnTyrLysGlySerThrSerTh 900  
Db 5415 -----CATCTCCAAAGTCAGGAGCGCCATTTTCTGAAAGCACTTCTAT 5461  
QY 900 rSerTyrGlySerAsnLeuAsnGlyLysIleProLeuThrPheGluAspLeuSerArgHi 920  
Db 5462 T-----GACAAATGCCCTGAGTCGACTGACCTTTGGGAATGAATCTCTGTCAACAA 5512  
QY 920 sGlnLeuHisAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlySe 940  
Db 5513 TGGGTACAGTCGAAGATTCAGATCTTTTCTGAACCTCCCTCCTGTGAT-----GAAA 5566  
QY 940 rLeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLy 960  
Db 5567 T-----GAAAGTTGGGCTTATCGCAGCGGACAAAACAGGTCCCG 5608  
QY 960 sLeuGlyValSerThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAl 980  
Db 5609 GTCTGCAATATCTATA-----TACAGACC 5632  
QY 980 aLeuAsnSerGlyMetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSe 1000  
Db 5633 TATCGACTATGGGATCTTTGGGAAGAACACACAGTTAGCTTTCTTACAAATCTAAGAG 5692  
QY 1000 rSerAlaAspPheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGl 1020  
Db 5693 GTCA-----CTTACAAAGAGAGATTTATGAAACCAAGTTTCTTAAGAACCCCTGG 5743  
QY 1020 yLysMetValHisProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAs 1040

Db 5744 C-----TTCTTAAAGATGATTG-----AGGAA 5767  
QY 1040 nPro 1041  
Db 5768 CCCT 5771

Search completed: May 17, 2004, 05:34:03  
Job time : 1320 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 17, 2004, 04:48:52 ; Search time 177 Seconds  
(without alignments)  
3314.029 Million cell updates/sec

Title: US-09-721-114-2

Perfect score: 5526

Sequence: 1 MEIVAVDQEGARVVGTCNML.....NKPADFTTISNDNEVMDYR 1057

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09721114/runat.14052004.092253.4135/app\_query.fasta\_1.1223  
-DB=Issued Patents NA -QMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blcsum62 -TRANS=human40.cdd  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09721114.qcgn 1.1.84 @runat.14052004.092253.4135 -NCPU=6 -ICPU=3  
-NO.WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLACK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELEX=7

Database : Issued Patents NA:

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCUTS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	298	5.4	8648	4	US-09-415-946-2
2	298	5.4	17341	4	US-09-415-946-1
3	214	3.9	4584	4	US-09-845-917A-51
4	214	3.9	4749	4	US-09-845-917A-55
5	214	3.9	5009	4	US-09-845-917A-52
6	214	3.9	5010	4	US-09-845-917A-56
7	214	3.9	5072	4	US-09-845-917A-2
8	214	3.9	5073	4	US-09-845-917A-1
9	214	3.9	9642	4	US-09-845-917A-30
10	214	3.9	10443	4	US-09-845-917A-25
11	214	3.9	13414	4	US-09-845-917A-27
12	211	3.8	7474	4	US-09-845-917A-26

#### ALIGNMENTS

##### RESULT 1

US-09-415-946-2

Sequence 2, Application US/09415946

Patent No. 6376751

GENERAL INFORMATION:

APPLICANT: Sung, Z. Renee

APPLICANT: Aubert, Dominique

APPLICANT: Chen, Lingjing

APPLICANT: The Regents of the University of California

TITLE OF INVENTION: Nucleic Acids That Control Reproductive Development in

TITLE OF INVENTION: Plants

FILE REFERENCE: 018941-00011005

CURRENT APPLICATION NUMBER: US/09/415,946

PRIOR FILING DATE: 1999-10-08

PRIOR FILING DATE: 1998-10-09

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 8648

TYPE: DNA

ORGANISM: Arabidopsis thaliana

FEATURE:

OTHER INFORMATION: EMP1 gene and its promoter region

NAME/KEY: promoter

LOCATION: (1)..(3201)

NAME/KEY: CDS

LOCATION: join(4241..4335, 4448..4623, 4704..4823, 4903..4956,

OTHER INFORMATION: EMBRYONIC FLOWER 1 (EMF1)

NAME/KEY: exon

LOCATION: (3202)..(3265)

OTHER INFORMATION: exon 1

NAME/KEY: intron

LOCATION: (3266)..(4159)

NAME/KEY: exon

LOCATION: (4160)..(4335)

Sequence 49, Appl  
Sequence 53, Appl  
Sequence 50, Appl  
Sequence 54, Appl  
Sequence 1626, Ap  
Sequence 1685, Ap  
Sequence 725, App  
Sequence 32, Appl  
Sequence 4, Appl  
Sequence 289, App  
Sequence 1051, Ap  
Sequence 1515, Ap  
Sequence 8, Appl  
Sequence 8, Appl  
Sequence 8, Appl  
Sequence 130, App  
Sequence 11, Appl  
Sequence 11, Appl  
Sequence 10, Appl  
Sequence 10, Appl  
Sequence 1, Appl  
Sequence 10, Appl  
Sequence 12, Appl  
Sequence 12, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 47, Appl  
Sequence 12, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl

US-09-845-917A-49  
US-09-845-917A-53  
US-09-845-917A-50  
US-09-845-917A-54  
US-09-134-001C-1626  
US-09-134-001C-1685  
US-09-376-594-725  
US-09-566-921-32  
US-08-931-999-4  
US-09-620-312D-289  
US-09-620-312D-1051  
US-09-134-001C-1515  
US-08-425-061-8  
US-08-825-886-8  
US-08-989-890-8  
US-09-620-312D-130  
US-08-425-061-11  
US-08-825-886-11  
US-08-989-890-11  
US-08-425-061-10  
US-08-825-886-10  
US-08-558-322-1  
US-08-989-890-10  
US-08-425-061-12  
US-08-825-886-12  
US-08-603-753D-1  
US-09-099-753-1  
US-08-986-106-1  
US-09-007-678B-47  
US-08-989-890-12  
US-08-480-784-1  
US-08-483-553-1  
US-08-487-002-1

205 3.7 4217 4  
205 3.7 4382 4  
205 3.7 4642 4  
205 3.7 4643 4  
179 3.6 6414 4  
197 3.6 3561 4  
196 3.6 8146 4  
196 3.6 11917 4  
196 3.6 6755 3  
196 3.5 6775 4  
188 3.4 7518 4  
188 3.4 2835 4  
184 3.5 3.3 5709 1  
184 3.5 3.3 5709 1  
184 3.5 3.3 5709 4  
184 3.5 3.3 8503 4  
183 3.3 5707 1  
183 3.3 5707 1  
183 3.3 5707 4  
183 3.3 5711 1  
183 3.3 5711 1  
183 3.3 5711 2  
183 3.3 5711 4  
183 3.3 5712 1  
183 3.3 5712 1  
183 3.3 5712 2  
183 3.3 5712 3  
183 3.3 5712 3  
183 3.3 5712 4  
183 3.3 5712 4  
183 3.3 5914 1  
183 3.3 5914 1  
183 3.3 5914 1



604 luValThrArggLyLysGlnThrMetIleAspAspIleProMetAspIleValGlnLeuL 624  
6536 AGACTACGACTGAGCAACAAACAC--GATGATATTCATGAGGATGATGAGCTCA 6592  
624 euAlaLysAsnGlnHisGluArgGlnLeuMet---ThrGluThrAspCysSerAspIleA 643  
6593 TGGCCAAAACACGATACGAGAGGTGCTTCCGCAACAAAGAGAAGATGTTAGCAACAAAC 6652  
643 snArgIleGlnSerLysThrAlaAspAspAspCysValIleValAlaAlaLysAspG 663  
6653 AGCATCACAGAAACACAGCACAAATCCAGAAATGCTTACTGATTTGATCTCAATGAAG 6712  
663 lySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGln---LysSerLeuA 681  
6713 CCTACGATAACGGGATCTCACTTGGAGCAACAAACACATCAAGACACCAACAAACCGTGA 6772  
681 laSerGlnSerThrGlnLysGlnLeuGlnGlyHisLeuAlaLeuThrThrGlnGlnSerP 701  
6773 GTAGCAACGCAAGAGGGAAGAA-----CATTTTCTTCTGGGAAGACACGACAACT 6823  
701 roHis-----ProGlnAsn-----PheGlns 708  
6824 CTCATGACTTCTTCCCAATAGTCAGCTTATGTGCTTCTCGTTTGGGATCTTCTCTC 6883  
708 erThrGlnGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerS 728  
6884 CTACCAAGAAACCGA-----GCCAGCT 6907  
728 erProLeuPheSerHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTrpG 748  
6908 CCATCCGGTTTTCTGGTCAACATCTGAGTGGCTTGGGAATTTGCCAATCTGTG-----G 6961  
748 lyArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSerP 768  
6962 GTAATCAGAACCTTCTCCATCTCATCTTCCGTTATTAACGTGCT----- 7005  
768 roAlaAlaThrCysGlyAla-----GlnPheArgProGlyIleGlnAlaValAspL 785  
7006 -TGATGATCTTCCAGAGTGTCTCTAATCAATACAGA----- 7041  
785 euThrSerThrHisValMetGlySerSerSerSerValAlaSerArgGlnProValIleA 805  
7042 --GAAGCTTCTCATCCATTTGGCCATCTCC-----ATGATAC 7078  
805 laProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProS 825  
7079 CACCACAGAGTCAATACAGCCAGTTTCTTTAAATATTATATCATGATCAACAAAT---CCGG 7135  
825 erThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValV 845  
7136 GTACGCTTTCA-----CAGGCACTCAACATGAAATATACATGACCTCACTTTGTTG 7189  
845 alLeuTyrProLysGlu----- 850  
7190 CTGCCAACGGGAGCAAAATGTGGGCTTAATCCGAATTTTCATTTGGCTGCACAAACATG 7249  
851 -----SerM 852  
7250 CTGCTGGGTTAGTAGTAGTAGTAGTAGGCAATAGATACCTTTCTAGTAGAGAGCTCTA 7309  
852 etProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPhePro----- 870  
7310 TACCGGATTCGATCTACTCAGCTTCTGATCTCTGCTGAGTCAACGACTCCCGCTG 7369  
871 --AsnTyrGlyThrSer-----SerArgAsnGlnMetGluSerGlnLeuHisAsnSerG 888  
7370 ACCAACCGGAAACATAAATTTACTAAAGACATTTTCCGCCAGGCAACCACTGCAAAAG 7429  
888 lnTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsnG 908  
7430 AGTTATAGAGCTTCAACACGGGAGCTCTAGTAGTACGCTACTCACTAAG----- 7482

908 lyLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgp 928  
7483 --CAGTACT-----TTTGATCTATACAGCAGAGATTCACACAGACCTTCCCGGA 7534  
928 roLeuArgPro---HisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIleA 947  
7535 AGAGTTTCCCATCACTCCACCTATTGGACG-----TCCTTCACTTTCATTTCAAAATG 7588  
947 laAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSerThrGlyI 967  
7589 CTTCAAGAGTCTCTCAT----- 7605  
967 leThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheS 987  
7606 -----CATCAAGAGAAGAAAACCAAGAGAAAACACACC-----TTTG 7642  
987 erAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAsp----- 1003  
7643 CTCCTGTTTACATACTCATGAAAACCGTGTTCAGACGACGACGACGACGACGACGACG 7702  
1004 -----PheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysM 1022  
7703 TCCAGCTGCTGGGAGCATCGAATTC----- 7729  
1022 etValHisProLeu-----AspArgPheValArgGlnAsp----- 1033  
7730 TGAATGCTTCTTGAATTTCAATGACGATTAAGAAAAGAAAAGAAAAGAAAAGACGAG 7789  
1034 ----- 1035  
7790 AGAGCTGCAATACAAATGCTCTGCGGACCTGTGAAGAACAGTTCTGACCCCATTTGT 7849  
1035 ylleThrAsnLysAsnProAlaAspPheThrThrIleSerAsnAspAsnGluTyrMet 1054  
7850 GCAGCGTCAATAGAAAACCTGCTGATTTCCACCATTCCTGGAACCTGGGAATGTTTACATG 7908

RESULT 2  
US-09-415-946-1  
; Sequence 1, Application US/09415946  
; Patent No. 6376751  
; GENERAL INFORMATION:  
; APPLICANT: Sung, Z. Renee  
; APPLICANT: Aubert, Dominique  
; APPLICANT: Chen, Lingjing  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Nucleic Acids That Control Reproductive Development in  
; TITLE OF INVENTION: Plants  
; FILL REFERENCE: 018941-000110US  
; CURRENT APPLICATION NUMBER: US/09/415,946  
; CURRENT FILING DATE: 1999-10-08  
; PRIOR FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 17341  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: genomic DNA (Ecotype Columbia) from CD82 clone  
; NAME/KEY: Promoter  
; LOCATION: (1)..(3201)  
; NAME/KEY: CDS  
; LOCATION: join(4241..4335, 4448..4623, 4704..4823, 4903..4956,  
; OTHER INFORMATION: EMBRYONIC FLOWER 1 (EMF1)  
; NAME/KEY: exon  
; LOCATION: (3202)..(3265)  
; OTHER INFORMATION: exon 1  
; NAME/KEY: intron  
; LOCATION: (3266)..(4159)  
; NAME/KEY: exon  
; LOCATION: (4160)..(4335)  
; OTHER INFORMATION: exon 2





QY 604 luValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuL 624  
 DB 6536 AGACTACGACTGAGAGCAAAACACAC---GATGATATTCCTCAATGAGATAGTAGGCTCA 6592  
 QY 624 euAlaLysAsnGlnHisGluArgGlnLeuMet---ThrGlnThrAspCysSerAspIleA 643  
 DB 6593 TGGCCAAAACCCAGTACGAGAGGTCTCTCCCGCAAAAGAGAAGATGTTAGCAAAAC 6652  
 QY 643 snArgIleGlnSerLysThrAlaAspAspCysValIleValAlaAlaLysAspG 663  
 DB 6653 AGCCATCACAGAAACAGCACACAAATCCAGATGCTCTACTGATTTGATCTCAATGAA 6712  
 QY 663 lysSerAspTyrAlaSerSerValPheAspThrAsnSerGlnGln-----LysSerLeuA 681  
 DB 6713 CCTACGATAACGGGATCTCACTTCGAGACAACACACATCAAGACCAACCAAAACCGGTGA 6772  
 QY 681 laSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGlnSerP 701  
 DB 6773 GTAGCAACGCAAGGAGGAGAA-----CATTTTCCTATCGGAGACAGCGCAACT 6823  
 QY 701 roHis-----ProGlnAsn-----PheGlns 708  
 DB 6824 CTCATGACTTCTCCCAATAAGTCAGCCTTATGTGCTTCTTCGTTTGGGATCTTTCCTC 6883  
 QY 708 erThrGlnGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerS 728  
 DB 6884 CTACCCCAAGAAACCGA-----GCCAGCT 6907  
 QY 728 erProLeuPheSerHisHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTrpG 748  
 DB 6908 CCATCCGGTTTCTCGTCACAACTGTCAGTGGCTTGGGAATTCGCCAAGCTG-----G 6961  
 QY 748 lyArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSerP 768  
 DB 6962 GTAATCAGAACCTTCTCCATCTCTCAATTCGGGTATTACGTGCT----- 7005  
 QY 768 roAlaAlaThrCysGlyAla-----GlnPheArgProGlyIleGlnAlaValAspL 785  
 DB 7006 --TGTGATCTTGCAGAGTGTCTCTATCATCAATACAGA----- 7041  
 QY 785 euThrSerThrHisValMetGlySerSerAsnTyrAlaSerArgGlnProValIleA 805  
 DB 7042 --GAAGCTTCTCATCAATTTGGCCATCTCC-----ATGATAC 7078  
 QY 805 laProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProS 825  
 DB 7079 CACCACAGAGTCAATACAGCCAGCTTCTTTAAATATTAAATCATGCTCAACAAAT---CCGG 7135  
 QY 825 erThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValV 845  
 DB 7136 GTACGCTTTCA-----CAGGCACTCAACATGAATATACATGAGACCTCAACTTGTG 7169  
 QY 845 alLeuTyrProLysGlu----- 850  
 DB 7190 CTGCCACGGGAAGCAAAATGTGGCCTAATCCAGAAATTTTCATTTGGCTGCACAAATG 7249  
 QY 851 -----SerM 852  
 DB 7250 CTGCTGGGTAGTAGTAGTAGTAGCCCAATAGATAACTTTCTAGTAGAGAGCTCTA 7309  
 QY 852 etProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPhePro----- 870  
 DB 7310 TACCGGATTCATCTACTACGCTCTCTGGATCTCTGGCTCTGAGGTCAACGACTCCCGCTG 7369  
 QY 871 --AsnTyrGlyThrSer-----SerArgAsnGlnMetGluSerGlnLeuHisAsnSerG 888  
 DB 7370 ACCAACCGGAACCAATAATTTACTAAAGACATTTTCCGCCAGCCCAACCGCTCCAAAG 7429  
 QY 888 lnTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsnG 908  
 DB 7430 AGTTTATAGAGCTTCAACAGGGAGTCTACTAGTACGCTCACTCAACTAAG----- 7482  
 QY 908 lyLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgP 928

DB 7483 --CAGATACCT-----TTTGATCTATACAGCAAGAGATTACACACAGCCTTCCCGGA 7534  
 QY 928 roLeuArgPro---HisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIleA 947  
 DB 7535 AGAGTTTCCCATCATCTCCACCTATTGGGAGC-----TCTTCACTTTCATTTCAAAATG 7598  
 QY 947 laAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSerThrGlyI 967  
 DB 7589 CTTTCATGGAGTCTCTCAT----- 7605  
 QY 967 leThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheS 987  
 DB 7606 -----CATCAAGAGAGAAACCAAGAGAGAAAGACACC-----TTTG 7642  
 QY 987 erAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAsp----- 1003  
 DB 7643 CTCCTGTTTACATACTCATGAAAAGCCGGTGTTCGACGACGACCAACGACCAAGCGAAGT 7702  
 QY 1004 -----PheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysM 1022  
 DB 7703 TCCAGCTGCTGGGAGCATCGAATTC----- 7729  
 QY 1022 etValHisProLeu-----AspArgPheValArgGlnAsp----- 1033  
 DB 7730 TGATGCTTCTTTGAAATTTTCATCATGACGATAAAGAAAGAAACAAAGAGAAAGCAG 7789  
 QY 1034 ----- 1035  
 DB 7790 AGAGCTGCAATAACAATCCCTCTCGCGGACCTCTGGAAGAACAGTCTTGGACCCCATTTGT 7849  
 QY 1035 ysIleThrAsnLysAsnProAlaAspPheThrIleSerAsnAspAsnGluTyrMet 1054  
 DB 7850 GCAGCGTCAATAGAACCCCTGCTGATTTCCACCATCTCTGAACTCGGAATGTTTACATG 7908

## RESULT 3

US-09-845-917A-51

; Sequence 51, Application US/09845917A

; Patent No. 6653529

; GENERAL INFORMATION:

; APPLICANT: Bogaert, Thierry

; APPLICANT: Vandekerckhove, Joel

; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH

; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND

; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR

; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR

; FILE REFERENCE: P/14-1

; CURRENT APPLICATION NUMBER: US/09/845,917A

; CURRENT FILING DATE: 2001-04-30

; NUMBER OF SEQ ID NOS: 58

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 51

; LENGTH: 4584

; TYPE: DNA

; ORGANISM: Caenorhabditis elegans

US-09-845-917A-51

## Alignment Scores:

Pred. No.:	5,37e-10	Length:	4584
Score:	214.00	Matches:	243
Percent Similarity:	34.58%	Conservative:	170
Best Local Similarity:	20.40%	Mismatches:	414
Query Match:	3.87%	Indels:	365
DB:	4	Gaps:	56

US-09-721-114-2 (1-1057) x US-09-845-917A-51 (1-4584)

QY 14 valGlyThrAsnCysMet-----LeuAlaArgGlyGlyThrGlyAla 27

DB 247 CTGGGTCTCGACTGCTCGAAACTCACCAAAACCGATATCGACGCGGAACCTTGGGTGCA 306

QY 28 valAlaProValLeuGluLeuThrAlaThrProArgGlnAspAlaAlaGluAlaGly 47

Db 307 GTTCTCCAGCTGCTCTCTCTCCCTCCACCTACAGCAG----- 345  
Qy 48 ValAspGluProAlaGlnHisGlnCysGluHisPheSerIleArgGlyTyrValAlaLeu 67  
Db 345 ----- 345  
Qy 68 LeuGlnLysAspProLysPheCysSerLeuSerArgIlePheHisAspGlnLysLys 87  
Db 346 -----AAGCTTCGGCAACTGAAAGAAAGATCAGAGAGAA 378  
Qy 88 CysAspGluHisLysAlaSerSerProPheSerValAlaLysPheArgArgTyrAsp 107  
Db 379 TTGAGCAACTACCCACATCCATTATGACCCCGGTTTCTAAATACCTCCGCCACGT 438  
Qy 108 CysSerLysCysLeuAspLysLeuLysThrSerAspAsnGlyThrAlaProArgThrLeu 127  
Db 439 GTCGCCACGTCAGCAACCGCTTCAGCAACTAACCCAAATTCACAAATTCACAAATGTC 498  
Qy 128 ProAlaLys---GlnAsnGlyThrSerAspGlyCysSerIleThrPheValArgSerThr 146  
Db 499 ACATCCAGGCTTCAGACTCCACAGTCAAGAAATTCGAAATTAATTCATCAAGATTTGGT 558  
Qy 147 PheValProAlaSerValGlySerGlnLysValSerProSerThrGlnSerSerGlnGly 166  
Db 559 ATCAAGCCAAAGAGCTCTGGACTTAACACCCCTCATCAACCACTTCATCAATAAT 618  
Qy 167 LysAsnAlaAspArgSerThrLeuProLysSerValGlnGluGlyAsnAsp----- 193  
Db 619 ACRAAATTCATTCGGT-----CCGTCGAGCGGTTTCGAGTGGCAATAATAATGTGGC 669  
Qy 184 -----SerLysCysAsnAlaProSerGlyLysAsnGlyAlaAlaGluAla 198  
Db 670 TCGACGATATCCACATCTCGAGAGGCTTAGAATCATCATCAACGTACAGCTCTAATTCG 729  
Qy 199 AsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAlaAla 218  
Db 730 AATCTAAACCGACCTACTCCCACTCCAAACCTTCTAGACCACAAACCCAGCTAGTT 789  
Qy 219 AsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIle 238  
Db 790 CGTGTCTCT-----ACAACTACAAATACTCGAAGC----- 819  
Qy 239 ThrTrpHisIleGluValAsnGlyAlaAspGlnProProSerThrProLysLeuSerGlu 258  
Db 820 -----TCAAAAGTAGCGCTCCGAAAGCCGTCGAGCACCCCAAACTTGTCTCT 867  
Qy 259 ValVal-----LeuLysArgAsnGluAsnGlyLysThrGluGluThrLeu 275  
Db 868 GTGAAGACTATTGGAGCAAAACAGACCCCGATACACGCGGCTGCTGTGTGTGGATG 927  
Qy 276 ValAlaGluGlnCysAsnLeuThrLysAspPro-----AsnProMetSer 290  
Db 928 CTGAATTAAGTTATTTCAGTAGCAAAACCCATCTTCTCATCGAATAGCCCAACCT 987  
Qy 291 GlyLysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLysProVal 310  
Db 988 ACAGAAAGCGCGCGGCTGCTCAACAACTTTGTCGAAATTCGTCGCCACGTG 1047  
Qy 311 ---SerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluValValLeuLys 329  
Db 1048 AAAAGTGGCTGAAG-----CGCGCAGCAGTAAGTGGGAAGT 1086  
Qy 330 ArgSerSerLysSerLys-----ArgLysThrAspLysLys 341  
Db 1087 GCCACGCTATGTCGAAGCTTTGTACGCCAAAGTTCTTACCGTAAACCGGACGCCCA 1146  
Qy 342 LeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAspAla 361  
Db 1147 ATCATATCTCAACAA-----GACTCG 1167  
Qy 362 LysLeuCys-ArgArgLysProLysLys-ValArgLeu-----LeuSerGluIleLea 379  
Db 1168 AAACGATGCTCAAAAGACGAGTGAAGAGAGTCCCGATACGTCGTGATTCACACAGCAGTC- 1226

Qy 379 snAlaAsnGlnValGluAspSerArg-SerAspGluValHisArgGluAsnAlaAlaAsp 398  
Db 1227 --GCCAAGCTCATCATCGACGAGAGGTTCCCTTAAGCATGCATTCACATCTTCCAGAGT 1284  
Qy 399 ProCysGluAspAspArgSerThrIleProValProMetGluValSerMetAspIle--- 417  
Db 1285 TCAACGTCAGAGAAAGTCTCCGTCATCAGACGATCTTACTCTTAAGCCCTCCATCGTG 1344  
Qy 418 -----ProValSerAsnHisThrValGlyGluAsp 427  
Db 1345 ACAGCTATCAGACAGCCGATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAGCT 1404  
Qy 428 -----GlyLeuLysSerSerLysAsnLysThrLysArgLysTyr 440  
Db 1405 GTTGAGGAAAAACCAACACTGCGCAGTGAAGAGTGAAGACAGCCGAAAGATCCCA 1464  
Qy 441 SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysArg 460  
Db 1465 CTTCCAGCTGTTCCGCCAGCTGACACC-----CAGCCCAACA 1500  
Qy 461 ThrGlySerValHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysVal 480  
Db 1501 ATCGGAGTTGTTAGTCCCAATTATGACAT-----AAGAGTTG 1539  
Qy 481 Thr-----ProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeu 498  
Db 1540 ACAATATGACCCGCTGATATCTGAAAAACACAGAACCTGAAAGCTCAATCAATGAGCATC 1599  
Qy 499 AspThrAsnMetHisLysThrAspVal--CysGlnHisValSerGluIleSerThrGln- 517  
Db 1600 GACACG-----ACGACGCTTCCACGCTTCCACCTCTAAATCAGTTGTTC 1647  
Qy 518 -----ArgCysSerSerLysGlyLysThr 526  
Db 1648 CTTAAATGACTTCAATCCGACACCAACACGTCAGTGTCTTCTTAAACACAGS--- 1703  
Qy 526 IaGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysTyrGlyGlyGluSerT 546  
Db 1704 -----AAAAATCAGTCGCTGTCAAG--TCGTTGATATGAGCAGT 1744  
Qy 546 hrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrG 566  
Db 1745 CG-----TCGCGCTCTG 1756  
Qy 566 LuAsnSerValLeuSerHis---SerAlaLysValSerPro-----AlaG 580  
Db 1757 AAGACTCATTGCTGCTCATGCTCGGCTCAGTCACTCCGCGCACAAAACTTCTGTGTA 1816  
Qy 580 LuHisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysL 600  
Db 1817 ATCATTCGCTG-----GAGAGAGGATGGGAAAGATTAAGCAT 1855  
Qy 600 ysGlnLysLeuGluValThr-----ArgGluL 609  
Db 1856 CAGATCCAGCGGCTACACTCTGACCGCGGTTGOGATGTGCGCAAAATGAGGAGA 1915  
Qy 609 ysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysAsnGlnH 629  
Db 1916 AGCTGAAGAATACGATGATGACTCGT-----CGAGCACGAAGCGCT 1960  
Qy 629 is-----GluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgI 645  
Db 1961 ATCTTGACAACTTCGAGACAGTCTCTCTCTGCTGCTGGAATATCCGATACAAACGAGC 2020  
Qy 645 leGlnSerLysThrThrAlaAsp-----AspAspCysValIleValAlaAlaLysA 662  
Db 2021 TCGACGACATATCCACGACGATTTGTCCGAGTAGACATGCGACAGTCGCTCCAAAC 2080  
Qy 662 spGlySerAspTyrAlaSerSerVal---PheAspThrAsnSerGlnGlnLys----- 678  
Db 2081 AT---AGGACTATTCCTGCTTGTTCGCCATCCCACTCTTCTTCTCTCAAGCCCCGAG 2137



499 ACATCCAGGCTTCAGACTCCACAGTCAAGAATATCGAAAATTGATTCATCAAAAGATTGGT 558  
147 PheValProAlaSerValGlySerGlnLysValSerProSerThrGlnSerSerGlnGly 166  
559 ATCAACCCAAAGACGCTCGACTTAACACCAACCCCTCATCATCAACCACTTCATCAAAATAT 618  
167 LysAsnAlaAspArgSerThrLeuProLysSerValGlnGlnGlyValAsnAsp 183  
619 ACAAAATTCATTCGGT-----CCGTGACGCGGTTCGAGTGGCAATTAATATGTTGGC 669  
184 -----SerLysCysAsnAlaProSerGlyLysAsnGlyAlaAlaGluAla 198  
670 TCGACGATATCCCATCTCGAAGACTTAGAATCATCATCAACGATACAGCTCTATTTCG 729  
199 AsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnLysrAspValAlaAla 218  
730 AATCTAAACGACCTACCTCCCAACTCCAAACCTTCTAGACCACCAACCCAGCTAGTT 789  
219 AsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIle 238  
790 CGGTGTTGCT-----ACAACTACAAAATCGGAAGC----- 819  
239 ThrTrpHisIleGluValAsnGlyAlaAspGlnProProSerThrProLysLeuSerGlu 258  
820 -----TCAAAGCTAGCCGCTCCGAAGCCGTGAGCACCCTCAAAATCTTCTTCT 867  
259 ValVal-----LeuLysArgAsnGluAspGluAsnGlyLysThrGluGluThrLeu 275  
868 GTGACGACTATTGGACAAAACAAGAGCCGATACACAGCGTGGTGGTGGTGGTGGTGGTGGT 927  
276 ValAlaGluGlnCysAsnLeuThrLysAspPro-----AsnProMetSer 290  
928 CTGAATTAAGTTATTTCAGTAGCAAAAACCCATCTTCCTCATCGAATAGCCCAACCT 987  
291 GlyLysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLysProVal 310  
988 ACGAGAAAGCGCGCGCGGTGGCTCAACAACTTCTCGAAATCGCTGCCCGCAGTG 1047  
311 -----SerGlyClnLysCysGluGlnIleCysAsnGluProCysGluGluValValLeuLys 329  
1048 AAAAGTGGCTGAAG-----CCGCCGACCACTAGCTAGCTGGAAGT 1086  
330 ArgSerSerLysSerLys-----ArgLysThrAspLysLys 341  
1087 GCCACGCTCTATGTCGAAGCTTTGTAGCCCAAAAGTTCTACCGTAAACGACGCCCA 1146  
342 LeuMetLysGlnGlnHisSerLysLysA-gThrAlaGlnAlaAspValSerAspAla 361  
1147 ATCATATCTCAACAA-----GACTCG 1167  
362 LysLeuCys-ArgArgLysProLysLys-ValArgLeu-----LeuSerGluIleIleA 379  
1168 AACGATGCTCAGAGAGCAGTAGAAGAGTCCGGATCGCTGGATTCAACAGCAGCTC- 1226  
379 snAlaAsnGlnValGluAspSerArg-SerAspGluValHisArgGluAsnAlaAlaAsp 398  
1227 --GCCAAGCTCATCATCGACGGAAGTTCCTTAAGCATGTCATTCACATCTTCCAAAGT 1284  
399 ProCysGluAspAspArgSerThrIleProValProMetGluValSerMetAspIle--- 417  
1285 TCACGTCGACGAGAAAGTCTCGTCATCAGCAGATCTTACTCTTAACGCCCTCCATCGTG 1344  
418 -----ProValSerAsnHisThrValGlyGluAsp 427  
1345 ACAGCTATCAGACGCGGATAGCGCAACACCGGTTTCTCCAAATATTATCAACAGCT 1404  
428 -----GlyLeuLysSerSerLysAsnLysThrLysArgLysLys 440  
1405 GTTGAGGAAAAACCAACTCGGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA 1464  
441 SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysArg 460  
1465 CTTCCAGCTGTTCCGCCACGTCGACACC-----CAGCCARCA 1500

461 ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
1501 ATCGAGGTGTAGTCCAAITATGGCAGAT-----AAGAGTTG 1539  
481 Thr-----ProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeu 498  
1540 ACAAAATGACCCGCTGATATCTGAAAACACAGACCTGAAAAGCTCCAAATCAATGACATC 1599  
499 AspThrAsnMetHisLysThrAspVal--CysGlnHisValSerGlnLysSerThrGln- 517  
1600 GACAGC-----ACGAGCTTCCACCGCTTCCACCTTCCAAATCAATGATTGTTCCA 1647  
518 -----ArgCysSerSerLysGlyLysThr 526  
1648 CTTAAAAGACTTCAATCCGACCAACCAACGATGTTCTTCTTAAACAAGG----- 1703  
526 laGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysLysLysLysLysLys 546  
1704 -----AAAAATCAGATCGCTGTCAG--TCGTTTGGATGATGACGAGT 1744  
546 hrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrG 566  
1745 CG-----TCCGCGTCTG 1756  
566 luAsnSerValLeuSerHis--SerAlaLysValSerPro-----Alag 580  
1757 AAGACTCCATTGTGCTCATCGCTCAGCTCAGTGGTACCTCCGCGCAGCAAAAACTTCTGCTGA 1816  
580 luHisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysL 600  
1817 ATCATTCGCTG-----GAGAGAGGATGGAAGAATAGACAT 1855  
600 ysGlnLysLeuGluValThr-----ArgGluL 609  
1856 CAGAACTCCAGCGGTACACCTCTGACGCGGTGTGGATGTGCGCCCAAAATGAGGGAGA 1915  
609 ysGlnThrMetIleAspAspIleProMetAspIleValGluLeuAlaLysAsnGlnH 629  
1916 AGCTGAAGAATACATGATGACTGCT-----CGAGCAGCAAGACGGCT 1960  
629 is-----GluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgI 645  
1961 ATCTCTGACAACTTCGAAACAGAGTTCCTCTGCTGCTGGAATATCCGATCAACACGAGC 2020  
645 leGlnSerLysThrThrAlaAsp-----AspAspCysValIleValAlaAlaLysA 662  
2021 TCGACGACATATCCAGGACGATTTGTCGGAGTAGACATGCGCAACAGCTGCGCTCCCAAC 2080  
662 spGlySerAspTyAlaSerSerVal---PheAspThrAsnSerGlnGlnLys----- 678  
2081 AT---AGCGACTATCCCACTTGTTCGCCATCCCACTCTCTCTCTCTCAAAAGCCCCAG 2137  
679 -----SerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlnGlnHisLeuAlaLeuThr 697  
2138 TCCCGAGTGGTCTCTCCACATCAGTCGATTCGATTCGAGCAGAGAA----- 2184  
697 hrGlnGluSerProHisProGlnAsnPheGlnSerThrGlnGlnGlnGlnHisLeuA 717  
2185 --CAGGAGAAATGTACAAACTCTGTCCTCCAGTCCGACGAGCAACGAGCCAGCTGCG- 2235  
717 rgMetGluGluMetValThrIleAlaAlaSerSerProLeuPheSerHis----- 734  
2236 -----CCCGTGGCAGCTCAACCTTCGACCAACATTCGCTAA 2272  
735 --AspAspGlnTyIleAlaGluAlaProThrGluHisTrpGlyArgLysAspAlaLysL 754  
2273 GATCCCGGATATCTATCTTATCTCA-----CACTATCA--GTGTCAGCTGATA 2323  
754 ysLeuThrTrpGluGlnPheLysAlaThrThrArg-----AsnSerProAlaAlaThrC 772  
2324 AGGACAAATGCTATGCTATGCTACAGACTAGTCAGCAGCTTCTTCTCAAAAACCAAGCT 2383

```

QY 772 ysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisValMetG 792
Db 2384 ATTGAGCCCAATTT-----CATTCACCTGAT-----CGTAAATGCCACTTCAAG 2428
QY 792 lySerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAspArgTyrAlaG 812
Db 2429 AGTTACATCCAGCAGCAGAGATGCGGCTCTCTTGAGCCG-----2472
QY 812 luArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGluA 832
Db 2473 -----AGACGGGTGCCGAACCTCGATGCGAAATATGATG 2506
QY 832 laSerLysLeuCysAspArgArgAsnAlaGly-----GlnValValLeuTyrProLysG 850
Db 2507 CTTGAGATCTACTCGCGGTTCCCGAGGTGGAAGCTCTACTGGTATCTATGGAGAG- 2565
QY 850 luSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPheP 870
Db 2566 -----ACGTTCCAACTGCACAGACTATCCGATGAAATAATCCCGGCACAT----- 2610
QY 870 roAsnTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerGlnTyrA 890
Db 2611 -----TCGCCAAAGTGAATGGATCCCACTA-----2640
QY 890 laHisAsnGlnTyrLysGlySerThrSerThrGlySerThrGlySerAsnLeuAsnGlyLysI 910
Db 2641 -----TCACCTGCTAGCAGCAGCAGCATATGATCT-----CTCAATGAGAAGT 2683
QY 910 leProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgProLeu- 929
Db 2684 AC-----GAACATGCTATTCCGGACATGGCAGCTGACTTGG 2719
QY 930 -----ArgProHisProArgValGlyValLeuG 939
Db 2720 AGTGTTACAGAACTGTCGACTCACTAACCAAGAACAGGAGAACTATGGAGCATGT 2779
QY 939 lySerLeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSer----- 957
Db 2780 TTGATCTTTTGGACAAAGCTTAGAAAACTCACTCAACACATTCGATCGCAACTTGA 2839
QY 958 -----GlyTyrLysLeuGlyValSerThrG 966
Db 2840 AGCCTGAAGGCGAATACGATTACGAGCAGCAGCATTCATTTGAGGGATATTAGCAATC 2899
QY 966 lyIleThrSerHisGlnMetAsnArgLysGlu-----976
Db 2900 ATCTTGATCCAACTCAGCTCATGCTAACGAAGCGCTGGTGAAGCTTCTTCGTCACACCAT 2959
QY 977 -----HisPheGluAlaLeuAsnSerGlyMetPheSerAlaLysT 990
Db 2960 CTCTGGAATCAGTTGCATCCCATCGATCATCGATGTCATCGTGCAGAAAGCAGCAAGC 3019
QY 990 rpAsnAlaLeuGlnLeuGlySerValSerSerAlaAspPheLeuSerAlaArgAsnS 1010
Db 3020 AGGAGAGATCAGCTGAGCTCGTTTGGCAGAAC-----3054
QY 1010 erIleAlaGlnSerTyrThrArgGlyLysGlyLysMetValHisProLeuAspArgPheV 1030
Db 3055 -----AAGAAAGCTGGATCCGCTCTCA-----CTCTCCAAAGTTCA 3091
QY 1030 alArgGlnAspIleCysIleThrAsnLysAsn-----ProAlaAspPheThrIleS 1048
Db 3092 CCAAGAG-----AAGNACAAGAACTACGACGAGCAGACATATGCCATCATTT 3139
QY 1048 erAsnAspAsnGluTyrMetAsp 1055
Db 3140 CCGGATCTCAAGAACTCTTGAC 3162

```

RESULT 5

US-09-845-917A-52

; Sequence 52, Application US/09845917A

; Patent No. 6653529

; GENERAL INFORMATION:

```

; APPLICANT: Bogaert, Thierry, Joel
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/845,917A
; CURRENT FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 5009
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
; US-09-845-917A-52

```

## Alignment Scores:

```

Pred. No.: 6,26e-10 Length: 5009
Score: 214.00 Matches: 243
Percent Similarity: 34.68% Conservative: 170
Best Local Similarity: 20.40% Mismatches: 414
Query Match: 3.87% Indels: 365
DB: 4 Gaps: 56

```

US-09-721-114-2 (1-1057) x US-09-845-917A-52 (1-5009)

```

QY 14 ValGlyThrAspCysMet-----LeuAlaArgGlyGlyThrGlyAla 27
Db 247 CTGGGTCTGAGCTGCTCGAAACTCACCACAAACCGATATCGACAGCGAAACTTGGGTGCA 306
QY 28 ValAlaProValLeuGluLeuThrAlaThrProArgGlnAspAlaAlaGluAlaGly 47
Db 307 GTTCTCCAGTGTCTTCTCTCTCCACTCAGCAG-----345
QY 48 ValAspGluProAlaGlnHisGlnCysGluHisPheSerIleArgGlyTyrValAlaIleu 67
Db 345 -----345
QY 68 LeuGlnLysLysAspProLysPheCysSerLeuSerArgIlePheHisAspGlnLysLys 87
Db 346 -----AAGCTTCGCACTGAAATAAGATCAGAAGAA 378
QY 88 CysAspGluHisLysAlaSerSerProPheSerValAlaLysPheArgTrpAsp 107
Db 379 TTGGAGCAACTCCCATCTCCATTATGCCCGGGTTCTAAATTTACCTCGCCACGT 438
QY 108 CysSerLysCysLeuAspLysLeuLysThrSerAspAsnGlyThrAlaProArgThrLeu 127
Db 439 GTGCGCACTGACCAACCGCTTCAGCAACTAACCCAAATTCACACTTCCACAAATGTCA 498
QY 128 ProAlaLys---GlnAsnGlyThrSerAspGlyCysSerIleThrPheValArgSerThr 146
Db 499 ACATCCAGGCTTCAGACTCCAGTCCAGATATCGAAATTTGATTCATCAAGATTGT 558
QY 147 PheValProAlaSerValGlySerGlnLysValSerProSerThrGlnSerSerGlnGly 166
Db 559 ATCAGCCCAAGACGCTGAGCTTAAACCCCTCATCATCAACCACTTCATCAATAAT 618
QY 167 LysAsnAlaAspArgSerThrLeuProLysSerValGlnGluGlyAsnAsp-----183
Db 619 ACAAAATTCATTCGT-----CGTCGAGCGGTTCGAGTGGCAATAATAATGTTGCG 669
QY 184 -----SerLysCysAsnAlaProSerGlyLysAsnGlyAlaGluAla 198
Db 670 TCGAGATATCCACATCTCGAAGAGCTTAGAATCATCATCAACGTCAGACTCTATTTCG 729
QY 199 AsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAlaAla 218
Db 730 AATCTAAACCGACTTACCTCCCACTCCAAACCTTCTAGACCACAAACCCAGCTAGTT 789
QY 219 AsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIle 238

```

Db 790 CGTGTTCCT-----ACAACTACAAAATCGAAGC----- 819  
 Qy 239 ThrTrpHisIleGluValAsnGlyAlaAspGlnProSerThrProLysLeuSerGlu 258  
 Db 820 -----TCAAAGCTAGCCGCTCCGAAAGCGTGAGCACCCCAAACTTGCTCT 867  
 Qy 259 ValVal-----LeuLysArgAsnGluAspGluAsnGlyLysThrGluGluThrLeu 275  
 Db 868 GTGAAGACTATTGGAGCAAAACAGAGCCCGGATAACAGCGGTGGTGGTGGTGAATG 927  
 Qy 276 ValAlaGluGlnCysAsnLeuThrLysAspPro-----AsnProMetSer 290  
 Db 928 CTGAATTAAGTTATTAGTACCAAAAACCCATCTCTCATCGAATAGCCCAACCT 987  
 Qy 291 GlyLysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProVal 310  
 Db 988 ACAGAAAGCGCGCGCGTGCCTCAACAACAACTTTGTGAAATCGCTCCCGCAGTG 1047  
 Qy 311 ---SerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluValValLeuLys 329  
 Db 1048 AAAAGTGGCTGAAG-----CCGCCGACAGTAAGCTGGAGT 1086  
 Qy 330 ArgSerSerLysSerLys-----ArgLysThrAspLysLys 341  
 Db 1087 GCCACGCTCTATGTCGAAGCTTTGTACGCCAAAGATTCTCTACCGTAAACAGGCGCCCA 1146  
 Qy 342 LeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGluAlaAspValSerAspAla 361  
 Db 1147 ATCATATCTCAACA-----GACTCG 1167  
 Qy 362 LysLeuCys-ArgArgLysProLysLys-ValArgLeu-----LeuSerGluIleLeuA 379  
 Db 1168 AAACGATGCTCAAAGAGCAGTGAAGAGATCGGATACGCTGGATTCAACAGCAGCTC- 1226  
 Qy 379 snAlaSerGlnValGluAspSerArg-SerAspGluValHisArgGluAsnAlaAsp 398  
 Db 1227 --GCCAACGTCAATCGACGGAAGTTCCTTAAGCATGATCTCCACATCTCTCCAGAGT 1284  
 Qy 399 ProCysGluAspAspArgSerThrIleProValProMetGluValSerMetAspIle--- 417  
 Db 1285 TCAACGTGACAGCAAAAGCTCGTCATCAGCAGATCTTACTTTAAACGCTCCATCGTG 1344  
 Qy 418 -----ProValSerAsnHisThrValGlyGluAsp 427  
 Db 1345 ACAGCTATCAGACGCGGATAGCCGCAACACCGGTTTCTCCAAATATTATCAACAGCT 1404  
 Qy 428 -----GlyLeuLysSerSerLysAsnLysThrLysArgLysTyr 440  
 Db 1405 GTTGAGGAAAACCAACACTGGCAGTGAAGAGTGAAGAGCACAGCGAAGAAAGATCCA 1464  
 Qy 441 SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysArg 460  
 Db 1465 CCTCCAGCTGTCCGCGCAGTGACACC-----CAGCCCAACA 1500  
 Qy 461 ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
 Db 1501 ATCGGAGTTGTAGTCAATATGCGCAT-----AAGAGTTG 1539  
 Qy 481 Thr-----ProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeu 498  
 Db 1540 ACAATGACCCCGTGATATCTGAAAACCAAGAACCTGAAAGCTCCAAATCAATGAGCATC 1599  
 Qy 499 AspThrAsnMetHisLysThrAspVal--CysGlnHisValSerGluIleSerThrGln- 517  
 Db 1600 GACACG-----ACGGACGTTCCACCGCTTCCACCTCTAAATCATGTTTCCA 1647  
 Qy 518 -----ArgCysSerSerLysGlyLysThrA 526  
 Db 1648 CTTAAAAATGACTTCAATCCGACCAACCAACCGTACGATGTTCTTCTAAACACAGG- 1703  
 Qy 526 laGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysTyrGlyGlyGluSerT 546  
 Db 1704 -----AAAAATCATCGCTGTCAAG--TCGTTTGGATATGAGCAGT 1744

Qy 546 hrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrG 566  
 Db 1745 CG-----TCCGCGCTGTG 1756  
 Qy 566 luAsnSerValLeuSerHis---SerAlaLysValSerPro-----AlaG 580  
 Db 1757 ARGATCCATTGTGCTCATGCTCGCTCAGGTAGCTCCGCCGCAAAAACCTTCTGTGTA 1816  
 Qy 580 luHisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysL 600  
 Db 1817 ATCAATTCGCTG-----GAGAGAAGGATGGGAAAGAAATAGACAT 1855  
 Qy 600 ysGlnLysLeuGluValThr-----ArgGluL 609  
 Db 1856 CAGAAATCCAGCGGTACACCTCTGACCGCGGTGTTCGATGTGCGCCAAATATGAGGGAGA 1915  
 Qy 609 ysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysAsnGlnH 629  
 Db 1916 AGCTGAAGATACGATGACATGACTGCT-----CGAGCACAGAAGCGCT 1960  
 Qy 629 is-----GluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgI 645  
 Db 1961 ATCCTGACAACTTCGAGACAGCTTCTCTGCTGCTGGAATATCCGATACACAGCAGC 2020  
 Qy 645 leGlnSerLysThrThrAlaAsp-----AspAspCysValIleValAlaAlaLysA 662  
 Db 2021 TCGACGACATATCCAGGACGATTTGTCGGAGTAGACATGCGCAACAGCTGCGCTCCAAC 2080  
 Qy 662 spGlySerAspTyrAlaSerSerVal---PheAspThrAsnSerGlnGlnLys----- 678  
 Db 2081 AT---AGCGACTATTCCCACTTGTTCGCCATCCACGCTCTTCTCTCAAAGCCCCGAG 2137  
 Qy 679 -----SerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThr 697  
 Db 2138 TCCCGAGTCGCTCCACATCGATTCGATTCGATTCGAGCAGAA----- 2184  
 Qy 697 hrGlnGluSerProHisProGlnAsnPheGlnSerThrGlnGlnGlnGlnThrHisLeuA 717  
 Db 2185 --CAGGAGAATGTGTACAAACTTCTGTCCAGTCGCAACGAGCAACGCTGCGC----- 2235  
 Qy 717 rgMetGluGluMetValThrIleAlaAlaSerSerProLeuPheSerHisHis----- 734  
 Db 2236 -----GCCGTGCCACTCCACCTTCGACCACTTCGCAACATTCGCTAA 2272  
 Qy 735 --AspAspGlnTyrIleAlaGluAlaProThrGluHisTrpGlyArgLysAspAlaLysL 754  
 Db 2273 GATCCCGGGATCTCATCTCTATTCTCA-----CACTTATCA---GTGTCAGCTGATA 2323  
 Qy 754 ysLeuThrTrpGluGlnPheLysAlaThrThrArg-----AsnSerProAlaAlaThrC 772  
 Db 2324 AGGACACAATCTCTATGCACTTCACAGACTAGTCAGACACTTCTTTCACAAAACCAAGCT 2383  
 Qy 772 ysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisValMetG 792  
 Db 2384 ATTCAGGCCAATTT-----CATTCACCTGAT---CGTAAATGCCACCTTCAAG 2428  
 Qy 792 lysSerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAspArgTyrAlaG 812  
 Db 2429 AGTTACATATCCACCGACGACAGAAATGCGGCTCTCTTGAGCCCG----- 2472  
 Qy 812 luArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGluA 832  
 Db 2473 -----AGACGGGTGCCGAACCTCGATGTCGAAATATGATT 2506  
 Qy 832 laSerLysLeuCysAspArgAsnAlaGly-----GlnValValLeuTyrProLysG 850  
 Db 2507 CTTGAGTCTACTCGCGCGTTCGCGAGTGAAGCTCTACTGTTATCTATGAGAG- 2565  
 Qy 850 luSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPheP 870  
 Db 2566 -----ACGTTCAACTGCACAGACTATCCGATGAATAATCCCGCACAT----- 2610





Db 1048 AAAAGTGCGCTGAAG-----CCGCCGACCAGTAAGCTGGGAAGT 1086  
Qy 330 ArgSerSerLysSerLys-----ArgLysThrAspLysLys 341  
Db 1087 GCCAGCTCTATCTGCAAGCTTTGTACGCCAAAAGTTTCTACCGTAAACAGCAGCCGCCA 1146  
Qy 342 LeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAspAla 361  
Db 1147 ATCATATCTCAACAA-----GACTCG 1167  
Qy 362 LysLeuLysArgArgLysProLysLysValArgLeu-----LeuSerGluLeuLeuLeu 379  
Db 1168 AAACGATGCTCAAGAGCAGTCAAGAGAGTCCGGATCGCTGGATTCAACAGCAGCTC- 1226  
Qy 379 snAlaAsnGlnValGluAspSerArg-SerAspGluValHisArgGluAsnAlaAlaAsp 398  
Db 1227 --GCCAAGCTCATCATCGAAGGAGTTCCTTAAGCATGTCATCCACATCTTCCAAAGAGT 1284  
Qy 399 ProCysGluAspArgSerThrIleProValProMetGluValSerMetAspIle--- 417  
Db 1285 TCAACGTCAGACGAAAGTCTCGTTCATCAGACGATCTTACTCTTAAGCGCTCCATCGTG 1344  
Qy 418 -----ProValSerAsnHisThrValGlyGluAsp 427  
Db 1345 ACAGCTATCAGACGCGATAGCCGCAACACCGGTTCTTCCAAATATATATCAACAGCGCT 1404  
Qy 428 -----GlyLeuLysSerSerLysAsnLysThrLysArgLysTyr 440  
Db 1405 GTTGAGGAAAAACCAACTGCGAGTGAAGAGTGAAGAGCAGCAGGAAAAAGATCCA 1464  
Qy 441 SerAspValValAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysArg 460  
Db 1465 CTTCCAGCTGTTCCGCGACGTGACACC-----CAGCCAA 1500  
Qy 461 ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
Db 1501 ATCGGAGTTGTAGTCCAAATATGACAT-----AAGAGTTG 1539  
Qy 481 Thr-----ProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeu 498  
Db 1540 ACAAATGACCCCGTGATATCTCAAAACCCAGAACCTGAAAGCTCCAATCAATGAGCATC 1599  
Qy 499 AspThrAsnMetHisLysThrAspVal--CysGlnHisValSerGluLeuSerThrGln- 517  
Db 1600 GACAGC-----ACGACGTTCCACCGCTTCCACCTCTAAATATCATGTTGTTCCA 1647  
Qy 518 -----ArgCysSerSerLysGlyLysThr 526  
Db 1648 CTATAAATGACTTCAATCCGACCAACCAACGATAGATGTTCTTCTTAAACAGG---- 1703  
Qy 526 laGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysTyrGlyGlyGluSerT 546  
Db 1704 -----AAAAATCATATCGCTCGCTCAAG--TCGTTTGGATATGACGAGT 1744  
Qy 546 hrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrG 566  
Db 1745 CG-----TCGCGGTCTG 1756  
Qy 566 luAsnSerValLeuSerHis---SerAlaLysValSerPro-----AlaG 580  
Db 1757 AAGACTCCATTGTGGCTCATGCTCGCTCAGGTGACTCCGCGCAAAATCTTCTGGTA 1816  
Qy 580 luHisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysL 600  
Db 1817 ATCAITTCGCTG-----GAGAGAGGATGGGAAGAATAAGACAT 1855  
Qy 600 ysglnLysLeuGluValThr-----ArgGluL 609  
Db 1856 CAGAATCCAGCGCTACACTCTGACGCGGTGTTGCGATGTGCGCCAAATATGAGGAGA 1915  
Qy 609 ysglnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysAsnGlnH 629  
Db 1916 AGCTGAAGAAATACGATGATGATCTGCT-----CGAGCAGCAGACGCGCT 1960

Qy 629 is-----GluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgI 645  
Db 1961 ATCTGTCAACTTCGAAGACAGTTCCTCTCTGTGCTGGATATCGATAACACAGC 2020  
Qy 645 leGlnSerLysThrThrAlaAsp-----AspAspCysValIleValAlaAlaLysA 662  
Db 2021 TCGACGACATATCCACGAGATTTGTCGGATAGACATGCGCAACAGTCGCTCCAAAC 2080  
Qy 662 spGlySerAspTyrAlaSerVal---PheAspThrAsnSerGlnGlnLys----- 678  
Db 2081 AT---AGCGACTATTCACACTTTGTCGCCATCCACAGTCTCTTCTCAAAGCCCCGAG 2137  
Qy 679 -----SerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThr 697  
Db 2138 TCCCGAGTCGCTCTCCACATCAGTCATCTCGATCTCGAGCAGAA----- 2184  
Qy 697 hrGlnGluSerProHisProGlnAsnPheGlnSerThrGlnGlnGlnGlnHisLeuA 717  
Db 2185 --CAGGAGAATGTGTACAAACTTCTGTCCAGTGCAGAACGAGCAACGTTGC----- 2235  
Qy 717 rgMetGluGluMetValThrIleAlaAlaSerSerProLeuPheSerHisHis----- 734  
Db 2236 -----GCCGCTGCCACCTCAACCTTCGGACACATTCGCTAA 2272  
Qy 735 --AspAspGlnTyrIleAlaGluAlaProThrGluHisTrpGlyArgLysAspAlaLysL 754  
Db 2273 GATCCCGGATCTCATCTTCTTCTCA-----CACTTATCA--GTGTCAGCTGATA 2323  
Qy 754 ysglnThrTrpGluGlnPheLysAlaThrThrArg-----AsnSerProAlaAlaThrC 772  
Db 2324 AGGACACATGTCTATGACTCAGACATGTCAGACCTTCTTCAAAAACCAAGCT 2383  
Qy 772 ysglyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisValMetG 792  
Db 2384 ATTCAGGCCCAATTT-----CATTCACTTGAT---CGTAAATGCCACCTTCAAG 2428  
Qy 792 lysSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAspArgTyrAlaG 812  
Db 2429 AGTTTCACATCCACGACGACAGATGCGGCTCTCTTGAGCCCG----- 2472  
Qy 812 luArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGluA 832  
Db 2473 -----AGCGGTGCGGAACTCGATCGTCAATATGATT 2506  
Qy 832 laSerLysLeuCysAspArgAsnAlaGly-----GlnValValLeuTyrProLysG 850  
Db 2507 CTTTCAGATCTACTCGCGGCTTCCGAGGTGGAAGCTCTACTGTATCTATGAGAG- 2565  
Qy 850 luSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPhe 870  
Db 2566 -----ACGTTCCAACTGCACGACTATCCGATGAATAATCCCCGACAT---- 2610  
Qy 870 roAsnTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerGlnTyrA 890  
Db 2611 -----TCTGCCAAAAGTCAGATGGATCCCACTA----- 2640  
Qy 890 laHisAsnGlnTyrLysGlySerThrSerTyrGlySerAsnLeuAsnGlyLysI 910  
Db 2641 -----TCACTGGTAGCAGCAGCATATGATCT--CTCAATGAGAGT 2683  
Qy 910 leProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgProLeu- 929  
Db 2684 AC-----GAACATGCTATTTCGGACATCGGACGTCGACTGG 2719  
Qy 930 -----ArgProHisProArgValGlyValLeuG 939  
Db 2720 AGTGTTCAGAACACTGTGCACTCACTAACAGAAACAGGAGAACTATGAGCACTGT 2779  
Qy 939 lysSerLeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSer---- 957  
Db 2780 TTGATCTTTTGACAAAGCTTAGAAAATCACTCAACACATGATGATCGATCCCACTGA 2839



1348	TCAAGCTCAGAGAAAGTCTCCGTCATCAGACGATCTTACTCTTAAAGCGCTCCATCTGGT	1407
418	-----ProValSerAsnHisThrValGlyGluAsp	427
1408	ACAGCTATCAGACAGCGGATAGCGCAACACCGGTTTCTCCAAATATTATCAACAAGCCT	1467
428	-----GlyLeuLysSerSerLysAsnLysThrLysAsnGlySerTyr	440
1468	GTTGAGGAAAAACCAACACTGGCAGTGAAGAGGAGTGAAGACAGCGAAAAAAGATCCA	1527
441	SerAspValValAspAspGlySerSerLysLeuMetAsnTrpLeuAsnGlyLysLysAsnG	460
1528	CCTCCAGCTGTCGCGCACTGACACC-----CAGCCAACA	1563
461	ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal	480
1564	ATCGAGTCTGTAGTCCAAATTATGGCACAT-----AAGAAGTTG	1602
481	Thr-----ProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeu	498
1603	ACAAATGACCCCGTGATATCTGAAAAACCAAGACCTGAAAAGTCCCATCAATGAGCATC	1662
499	AspThrAsnMetHisLysThrAspVal-----CysGlnHisValSerGluIleSerThrGln	517
1663	GACACG-----ACGAGTCTCCACCGCTTCCACCTCTCTAAATCAGTTGTTCCTCA	1710
518	-----ArgCysSerSerLysGlyLysThr	526
1711	CTTAAATAGACTTCAATCCGACACACCACCGTACGATGTTCTCTTAAACCAAGG---	1766
526	laGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysTyrGlyGlyGluSerT	546
1767	-----AAAAATCACATCGCTGTCAAG--TCGTTTGGATATGAGCAGT	1807
546	hrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrG	566
1808	CG-----TCGCGGCTGT	1819
566	luAsnSerValLeuSerHis-----SerAlaLysValSerPro-----AlaG	580
1820	AAGACTCATTTGCGCTCATCGCTCGCTCAGGTGACTTCGCGCACAAAACTCTCGTA	1879
580	luHisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysL	600
1880	ATCATTCGCTG-----GAGAGAAGCATGGGAAAGAATAAGACAT	1918
600	ysGlnLysLeuGluValThr-----ArgGluL	609
1919	CAGAATCAGCGGCTACACCTCTCAGCGCGGTGTTCGATGTGCGCCAAATAGGGAGA	1978
609	ysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysAsnGlnH	629
1979	AGCTGAAAGAATACGATGACATGACTCGT-----CGAGCACAGAAACGGCT	2023
629	is-----GluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgI	645
2024	ATCTCGACAACTTCGAGACAGACTTCTCTTGTGCTTCGTAATATCGATAACACGAGC	2083
645	leGlnSerLysThrThrAlaAsp-----AspAspCysValIleValAlaAlaLysA	662
2084	TCGAGCATATCCACGACGATTTGTCGAGTAGACATGGCAACACAGTCCCTCCAAAC	2143
662	spGlySerAspTyrAlaSerSerVal---PheAspThrAsnSerGlnGlnLys-----	678
2144	AT--AGCGCATATTCCCACTTTGTTGGCCATCCCACTCTCTCTCTCTCAAAAGCCCGAG	2200
679	-----SerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlnHisLeuAlaLeuThrT	697
2201	TCCCGAGTCCGCTCTCCACATGCTCGATTCTCGATCTCGAGCAGAA-----	2247
697	hrGlnGluSerProHisProGlnAsnPhenGlnSerThrGlnGlnGlnThrHisLeuA	717
2248	--CAGGAGAAAGTGTACAAACTTCTGCTCCAGTCCGACGAGCAACGCTGC-----	2298

Qy	717	rgMetGluGluMetValThrIleAlaAlaSerSerProLeuPheSerHisHis	-----	734
Db	2299	-----GC	CGCGTGCACCTTCAACCTTCGGACAACTTCGCTAA	2335
Qy	735	--AspAapGlnTyrIleAlaGluAlaProThrGluHisTrpGlyArgLysAspAlaLysL	754	
Db	2336	GATCCCGGATACTCATCTATTCTCCA	-----CACTTATCA	2386
Qy	754	ysLeuThrTrpGluGlnPheLysAlaThrArg	-----AsnSerProAlaAlaThrC	772
Db	2387	AGGACAAATGCTATGCACATCAACAGTAGTCGACGACCTTCTTC	CAAAAACCAAGCT	2446
Qy	772	ysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisValMetG	792	
Db	2447	ATTTCAGGCCAATT	-----CAUTCACCTTGAT	2491
Qy	792	lySerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuLeuAspArgTyrAlaG	812	
Db	2492	AGTTTCACCTCCACCGACACAGAAATCGCGCTCTCTTGAGCCCG	-----	2535
Qy	812	luArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGluA	832	
Db	2536	-----AGACGGGTGCGGAAC	TGATATGAT	2569
Qy	832	laSerLysLeuCysAspArgArgAsnAlaGly	-----GlnValValLeuTyrProLysG	850
Db	2570	CTTCAGGATCCTACTCGCGCGTTC	CCGAGTGGAAAGCTCTACTGCTATCTATGGAGAG	2628
Qy	850	luSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPheP	870	
Db	2629	-----ACGTTCCAACTGCACAGACTATCCGATGAAAAATCCCGCACAT	-----	2673
Qy	870	rolanTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerGlnTyrA	890	
Db	2674	-----TCGCCAAAGTGAGATGGATGCCCACTA	-----	2703
Qy	890	laHisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsnGlyLysI	910	
Db	2704	-----TCAC	TGCTAGCAGCAGACATATGATCT	2746
Qy	910	leProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisAspProLeu	929	
Db	2747	AC	-----GAACATGCTATTCGGACATCGGCACGTGACTTGG	2782
Qy	930	-----ArgProHisProArgValGlyValValLeuG	939	
Db	2783	AGTGTTACAGAACACTGTCGACTCACTAACCAAGAACAGGAGAACTATGGAGCATTTGT	2842	
Qy	939	lySerLeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSer	957	
Db	2843	TTGATCTTTTGGCAAAAGCTTAGAAAACTCACTCAACATTCATCGATCCCACTTGA	2902	
Qy	958	-----GlyTyrLysLeuGlyValSerThrG	966	
Db	2903	AGCCTGAAGAGCAATACGATTCAGGCAGGACATTCATTCATTTGAGGGATATTAGCAATC	2962	
Qy	966	lyIleThrSerHisGlnMetAsnArgLysGlu	-----	976
Db	2963	ATCTTGATCCAACTACCTAGCTCATGCTAACGAGGCGCTGGTGAGCTTCTTCGTCAACCAT	3022	
Qy	977	-----HisPheGluAlaLeuAsnSerGlyMetPheSerAlaIlyst	990	
Db	3023	CTCTCGAATCAGTTGCATCCATCGATCATCGATGTCATCGTCTG	CAAAAGCAGCAAGC	3082
Qy	990	rpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAspPheLeuSerAlaArgAsnS	1010	
Db	3083	AGGAGAAGATCAGCTTGAGCTCGTTGGCAAGAAC	-----	3117
Qy	1010	erIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHisProLeuAspArgPheV	1030	
Db	3118	-----AAGACAGCTTGATCGCTCTCA	-----CTCTCCAAGTCA	3154



Db 1603 ACAATGACCCCGTGATATCTGAAACACAGAACCTGAAAGCTCCAAATCAATGAGCATC 1662  
Qy 499 AspThrAsnMetHisLysThrAspVal--CysGlnHisValSerGluIleSerThrGln- 517  
Db 1663 GACACG-----ACGAGCTTCCACCGCTTCCACCTCTAAATCATGTTTCCA 1710  
Qy 518 -----ArgCysSerSerLysGlyLysThr 526  
Db 1711 CTTAAATGACTTCAATCCGACACACCAACGATGATGTTCTCTTAAACAAAG----- 1766  
Qy 526 laGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysThrGlyGlyGluSerT 546  
Db 1767 -----AAAATCATCGCTGTCAAG--TCGTTGATATGAGCAGT 1807  
Qy 546 hrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrG 566  
Db 1808 CG-----TCGCGGTCTG 1819  
Qy 566 luAsnSerValLeuSerHis--SerAlaLysValSerPro-----AlaG 580  
Db 1820 AGACTCCCAATGGCTCATGCTCGCTCGCTGACTCGCGACAAAACCTTCTGGTA 1879  
Qy 580 luHisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysL 600  
Db 1880 ATCATTCGCTG-----GAGAGAAGGATCGGAAAGAAATAGACAT 1918  
Qy 600 ysGlnLysLeuGluValThr-----ArgGluL 609  
Db 1919 CAGAAATCCAGCGGCTACACTCTGACGCGGTGTTGCGATGTCGCGCAAAATGAGGAGA 1978  
Qy 609 ysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysAsnGlnH 629  
Db 1979 AGTGAAGATACGATGACATGACTCGT-----CGAGCACAGACGGCT 2023  
Qy 629 is-----GluA--glnLeuMetThrGluThrAspCysSerAspIleAsnArgI 645  
Db 2024 ATCTGCACAACTTCGAAGACAGTCTCTCTGCTCGAATATCCGATAACACAGC 2083  
Qy 645 leGlnSerLysThrThrAlaAsp-----AspAspCysValIleValAlaAlaLysA 662  
Db 2084 TCAGACACATATCCAGGAGATTTGTCGGAGTAGACATGCGACAGTCGCTCCAAAC 2143  
Qy 662 spGlySerAspTyrAlaSerSerVal---PheAspThrAsnSerGlnGlnLys----- 678  
Db 2144 AT---AGCGACTATCCCACTTGTTCGCCATCCACGCTCTCTCTCTCAAAGCCCGAG 2200  
Qy 679 -----SerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrT 697  
Db 2201 TCCCGAGTCGGTCTCCACATCATCGATCTCGATCTCGACGAA----- 2247  
Qy 697 hrGlnGluSerProHisProGlnAsnPheGlnSerThrGlnGluGlnGlnThrHisLeuA 717  
Db 2248 --CAGGAGATGTGTACAAACTTCTGTCCAGTCCGACGCGAACGACCAACGTCGC----- 2298  
Qy 717 rgMetGluGluMetValThrIleAlaAlaSerSerProLeuPheSerHisHis----- 734  
Db 2299 -----GCGCTGCCACCTCAACCTTCGGACACATTCGCTAA 2335  
Qy 735 --AspAspGlnTyrIleAlaGluAlaProThrGluHisThrGlyArgLysAspAlaLysL 754  
Db 2336 GATCCCGGGATACATCATCTATCTCA-----CACTTATCA---GTGTCAGCTGATA 2386  
Qy 754 ysLeuThrTrpGluGlnPheLysAlaThrThrArg-----AsnSerProAlaAlaThrC 772  
Db 2387 AGGACACAATGTCTATGCACTACAGACTAGTCGACGACCTCTTCACAAAAACCAAGCT 2446  
Qy 772 ysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisValMetG 792  
Db 2447 ATTCAAGCCCAATT-----CATTCATTGAT---CGTAATGCCACTTTCAG 2491  
Qy 792 lysSerSerAsnThrAlaSerArgGlnProValIleAlaProLeuAspArgTyrAlaG 812  
Db 2492 AGTTCATCCACCGACAGCAATGCGGCTCTCTTGAGCCCG----- 2535

Qy 812 luArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGluA 832  
Db 2536 -----AGACGGGTGCGCACTCGATGTCGAAATATGATT 2569  
Qy 832 laSerLysLeuCysAspArgArgAsnAlaGly-----GlnValValLeuTyrProLysG 850  
Db 2570 CTTCAAGATCTACTCGCGGTTCCCGAGGTGGAAGCTCTACTGTTATCTATGAGAG- 2628  
Qy 850 luSerMetProAlaThrHisIleLeuArgMetMetAspProSerThrLeuAlaSerPheP 870  
Db 2629 -----ACGTTCCAACTGCACAGACTATCCGATGAAAAATCCCCCACAAT----- 2673  
Qy 870 roAsnTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerGlnTyrA 890  
Db 2674 -----TCGCCAAAAGTGGATGGATCCCACTA----- 2703  
Qy 890 laHisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsnGlyLysI 910  
Db 2704 -----TCACTGGCTAGCACACAGCATATGGATCT---CTCAATGAGAAGT 2746  
Qy 910 leProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgProLeu- 929  
Db 2747 AC-----GAACATGCTATTTCGGGACATGSCACGTGACTGG 2782  
Qy 930 -----ArgProHisProArgValGlyValLeuG 939  
Db 2783 AGTGTTCACAGACACACTGCTGACTCACTAACCAAGAACAGAGAACTATGGAGCATGT 2842  
Qy 939 lysSerLeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSer---- 957  
Db 2843 TTGATCTTTTGGACAAAGCTTAGAAACTCACTACACATTCATGATCGATCCAACTGA 2902  
Qy 958 -----GlyTyrLysLeuGlyValSerThrG 966  
Db 2903 AGCCTGAAGGCAATACGATTCAGGAGGACATGCTCATTTGAGGATATTAGCAATC 2962  
Qy 966 lyIleThrSerHisGlnMetAsnArgLysGlu----- 976  
Db 2963 ATCTTGCACTCAACTCAGCTCAGCTACAGAAAGCGCTGGTGAGCTTCTCGTCAACCAT 3022  
Qy 977 -----HisPheGluAlaLeuAsnSerGlyMetPheSerAlaLysT 990  
Db 3023 CTCGTGAATCAGTTCATCCATCGATCATCGATGTCATCTCGTCGAAAAGCAGCAAGC 3082  
Qy 990 rpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAspPheLeuSerAlaArgAsnS 1010  
Db 3083 AGGAGAAGATCAGCTTGAGCTCGTTTGGCAAGAAC----- 3117  
Qy 1010 erIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHisProLeuAspArgPheV 1030  
Db 3118 -----AAGAGAGCTGGATCCGCTCCTCA-----CTCTCCAAGTTCA 3154  
Qy 1030 alaArgGlnAspIleCysIleThrAsnLysAsn-----ProAlaAspPheThrThrIleS 1048  
Db 3155 CCAAGAAAG-----AAGAACAAAGAACTACAGCAAGACATATGCTCATCAATTT 3202  
Qy 1048 erAsnAspAsnGluTyrMetAsp 1055  
Db 3203 CCGGATCTCAAGGAAGACTCTTGAC 3225

## RESULT 9

US-09-845-917A-30  
; Sequence 30, Application US/09845917A  
; Patent No. 653529  
; GENERAL INFORMATION:  
; APPLICANT: Bogaert, Thierry  
; APPLICANT: Vandekerckhove, Joël  
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH  
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND  
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR  
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR  
; FILE REFERENCE: P/14-1

; CURRENT APPLICATION NUMBER: US/09/845,917A  
 ; CURRENT FILING DATE: 2001-04-30  
 ; NUMBER OF SEQ ID NOS: 58  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 30  
 ; LENGTH: 9642  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: artificial  
 ; OTHER INFORMATION: plasmid  
 US-09-845-917A-30

# Alignment Scores:

Pred. No.: 1.95e-09 Length: 9642  
 Score: 214.00 Matches: 243  
 Percent Similarity: 34.68% Conservative: 170  
 Best Local Similarity: 20.40% Mismatches: 414  
 Query Match: 3.87% Indels: 365  
 DB: 4 Gaps: 56

## US-09-721-114-2 (1-1057) x US-09-845-917A-30 (1-9642)

QY 14 ValGlyThrAsnCysMet-----LeuAlaArgGlyThrGlyVala 27  
 DB 1337 CTGGGTCGACTGCTCGAACTCACCAGAACCGATATCGACGCGGAACTTGGGTGCA 1396  
 QY 28 ValAlaProValLeuLeuThrAlaThrProArgGlnAspAlaAlaGluAlaGly 47  
 DB 1397 GTTCTCCAGCTGCTTCTCTGCTCTCCACCTACAGCAG-----1435  
 QY 48 ValAspGluProAlaGlnHisGlnCysGluHisPheSerIleArgGlyTyrValAlaLeu 67  
 DB 1435 -----1435  
 QY 68 LeuGlnLysLysAspProLysPheCysSerLeuSerArgIlePheHisAspGlnLysLys 87  
 DB 1436 -----AAGCTTCGGCAACTCGAAAGAGATCAGAAAGAA 1468  
 QY 88 CysAspGluHisLysAlaSerSerProPheSerValAlaLysPheArgGlyTyrAsp 107  
 DB 1469 TTGGAGCAACTACCCATCATTCACCCCGGTTTCTAAATACCTCGCCACGT 1528  
 QY 108 CysSerLysCysLeuAspLysLeuLysThrSerAspAsnGlyThrAlaProArgThrLeu 127  
 DB 1529 GTCCCAAGCTCAGCAACCGCTTCAGCACTAACCCAAATTCCAACTTTCACCAATGTC 1588  
 QY 128 ProAlaLys---GlnAsnGlyThrSerAspGlyCysSerIleThrPheValArgSerThr 146  
 DB 1589 ACATCCAGGCTTCAGACTCCAGACTCAGCAATATCGAAATTCATTCATCAAGATGGT 1648  
 QY 147 PheValProAlaSerValGlySerGlnLysValSerProSerThrGlnSerSerGlnGly 166  
 DB 1649 ATCAAGCAAGAGCTCTGACTTAACACCCCTCATCATCACTTTCATCAATAAT 1708  
 QY 167 LysAsnAlaAspArgSerThrLeuProLysSerValGlnGluGlyAsnAsp-----183  
 DB 1709 ACAAAATTCATTCCT-----CCGTGAGCGGTTCGAGTGCATAATAATATGTTGGC 1759  
 QY 184 -----SerLysCysAsnAlaProSerGlyLysAsnGlyAlaAlaGluAla 198  
 DB 1760 TCGACGATATCCCATTCGGAAGAGCTTAGAATCATCATCATCATCATCATCATCATTCG 1819  
 QY 199 AsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnThrAspValAlaAla 218  
 DB 1820 AATCTAAACCGACTTACTCCCACTCCAAAACCTTCTAGACCAACACCCAGCTAGTT 1879  
 QY 219 AsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIle 238  
 DB 1880 CGTGTTCCT-----ACAACTACAAAATCGGAAGC-----1909  
 QY 239 ThrTrpHisIleGluValAsnGlyValaAspGlnProProSerThrProLysLeuSerGlu 258  
 DB 1909 -----1909

DB 1910 -----TCAAGCTAGCCGCTCGAAAGCGGTGAGCAGCCCAAACTGTCTTCT 1957  
 QY 259 ValVal-----LeuLysArgAsnGluAspGluAsnGlyLysThrGluGluThrLeu 275  
 DB 1958 GTGAAGACTATTGGAGCAAAACAGAGCCGCAATACAGCGGTGGTGGTGGTGAATG 2017  
 QY 276 ValAlaGluGlnCysAsnLeuThrLysAspPro-----AsnProMetSer 290  
 DB 2018 CTGAATTAAGTTATTTCAGTAGCAAAACCCATCTTCTCTCATCGAATAGCCCAACCT 2077  
 QY 291 GlyLysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLysProVal 310  
 DB 2078 ACGAGAAGCGCGCGCGTCTCAACAACAACCTTGTGGAATATCGTCCGCCAGTG 2137  
 QY 311 ----SerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluValValLeuLys 329  
 DB 2138 AAAAGTGGCTGAAG-----CCGCGAGCCAGTAGTGGGAAGT 2176  
 QY 330 ArgSerSerLysSerLys-----ArgLysThrAspLysLys 341  
 DB 2177 GCCACGTCTATGTGGAAGCTTTGTAGCCCAAGTTTCTACCGTAACAGCGAGCCCA 2236  
 QY 342 LeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAspAla 361  
 DB 2237 ATCATATCTCAACA-----GACTCG 2257  
 QY 362 LysLeuCys-ArgArgLysProLysLys-ValArgLeu-----LeuSerGluIleAla 379  
 DB 2258 AAACGATGCTCAAGAGAGAGTGAAGAGAGTCCGATACGCTGGATTCACAGCAGTC- 2316  
 QY 379 snAlaAsnGlnValGluAspSerArg-SerAspGluValHisArgGluAsnAlaAlaAsp 398  
 DB 2317 -GCCAACGTCTATCGAGCGGAAGGTTCCCTAAGCATGCATTCACATCTTCCAGAGT 2374  
 QY 399 ProCysGluAspAspArgSerThrIleProValProMetGluValSerMetAspIle--- 417  
 DB 2375 TCAACGTGACAGAAAGTCTCGTCAACAGAGATCTTACTCTTAAACGCTCCATCGTG 2434  
 QY 418 -----ProValSerAsnHisThrValGlyGluAsp 427  
 DB 2435 ACAGCTATCAGACGCGGATAGCCGCAACACCGGTTTCTCAAAATATTATCAACAGCT 2494  
 QY 428 -----GlyLeuLysSerSerLysAsnLysThrLysArgLysTyr 440  
 DB 2495 GTTAGCAAAACCAACACTGCGCAGTGAAGAGGTGAAGAGCAGCAGCAAAAGATCCA 2554  
 QY 441 SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysArg 460  
 DB 2555 CTCCAGCTGTTCCGCCAGCTGCACAC-----CAGCCCAACA 2590  
 QY 461 ThrGlySerValHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
 DB 2591 ATCGGAGTTGTAGTCCAATTATGGCAGCAT-----AAGAGATTG 2629  
 QY 481 Thr-----ProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeu 498  
 DB 2630 ACAATGACCCCGTGATATCTGAAGAACCGAGAGCTGAAGAGCTCCATCATCATCATC 2689  
 QY 499 AspThrAsnMetHisLysThrAspVal--CysGlnHisValSerGluIleSerThrGln- 517  
 DB 2690 GACACG-----ACGAGCGTTCCACCGCTTCCACCTTCAAAATCAGTTGTTC 2737  
 QY 518 -----ArgCysSerSerLysGlyLysThr 526  
 DB 2738 CTTAAATGACTTCAATCCGACCAACCAACGATAGTGTCTTCTTAAACAGG----- 2793  
 QY 526 laGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysTyrGlyGlyGluSer 546  
 DB 2794 -----AAAAATCATCGCTGTCAG--TCGTTTGGATATGACAGT 2834  
 QY 546 hrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrG 566  
 DB 2835 CG-----TCGCGCTGTG 2846

Qy	890	laHiSaSnGlnTyrLysGlySerThrSerTyrGlySerAlaLeuAsnGlyLysI	910
Db	3731	-----TCATCGGCTAGCACACACATATCGATCT-----CTCAATGAGAACT	3773
Qy	910	leProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgProLeu	929
Db	3774	AC-----GACATGCTATTTCGGACATCGGCACGCTGACTTGG	3809
Qy	930	-----ArgProHisProArgValGlyValLeuG	939
Db	3810	AGTGTTTACAAGAACACTGTCGACTCACTAACCAAGAAACAGGAGAACTATGGAGCATTTCT	3869
Qy	939	lySerLeuLeuGlnLysGluIleAlaAsnTrpSerClnuAsnCysGlyThrGlnSer	957
Db	3870	TTGATCTTTTGGACAAAAGCTTAGAAAACCTCACTCAACACATTTGATTCGATCCAACTTGA	3929
Qy	958	-----GlyTyrLysLeuGlyValSerThrG	966
Db	3930	AGCCTGAAGAGGCAATACGATTCAAGCAGGACATTGCTCATTTTGGGGATATTAGCAATC	3989
Qy	966	lyIleThrSerHisGlnMetAsnArgLysGlu	976
Db	3990	ATCTTCATCCAACTCAGCTCATGCTTAACGAAAGCGCTCGTGAGCTTCTTCGTCAACCAT	4049
Qy	977	-----HisPheGluAlaLeuAsnSerGlyMetPheSerAlaLysT	990
Db	4050	CTCTGGAATCAGTTGCATCCCATCGATCATCGATGTCATCGTCGTCGTAAGAAAGCAGCAAGC	4109
Qy	990	rPaSnAlaLeuGlnLeuGlySerValSerSerAlaAspPheLeuSerAlaArgAsnS	1010
Db	4110	AGGAGAAAGATCAGCTTCGCTCGTTGGCAAGAAC	4144
Qy	1010	erIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHisProLeuAspArgPheV	1030
Db	4145	-----AAGAAGAGCTCGATCCGCTCTCTCA	4181
Qy	1030	alArgGlnAspIleCysIleThrAsnLysAsn-----ProLaaSpPheThrThrIleS	1048
Db	4182	CCAAGAAG-----AAGAACACAGAACTAGACGACGACATATGCCATCAATTT	4229
Qy	1048	erAsnAspAsnGluTyrMetAsp	1055
Db	4230	CCGATCTCAAGAACTCTTGAC	4252

RESULT 10  
 US-09-845-917A-25  
 ; Sequence 25, Application US/09845917A  
 ; Patent No. 6653529  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bogaert, Thierry  
 ; APPLICANT: Vandekerckhove, Joel  
 ; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH  
 ; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND  
 ; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR  
 ; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR  
 ; FILE REFERENCE: P/14-1  
 ; CURRENT APPLICATION NUMBER: US/09/845,917A  
 ; CURRENT FILING DATE: 2001-04-30  
 ; NUMBER OF SEQ ID NOS: 58  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 25  
 ; LENGTH: 10443  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: artificial  
 ; OTHER INFORMATION: plasmid  
 US-09-845-917A-25

Alignment Scores:		
Pred. No.:	2,24e-09	Length: 10443
Score:	214.00	Matches: 243
Percent Similarity:	34.68%	Conservative: 170

Alignment Scores:	
Pred. No.:	2 24e-09
Score:	214.00
Percent Similarity:	34.68%
Matches:	243
Conservative:	170
Length:	10443

Best Local Similarity: 20.40%		Mismatches: 414	
Query Match:	3.87%	Indels:	365
DB:	4	Gaps:	56
US-09-721-114-2 (1-1057) x US-09-845-917A-25 (1-10443)			
Qy	14	ValGlyThrAsnCysMet-----	LeuAlaArgGlyGlyThrGlyAla 27
Db	257	CTGGGTCTCGACTGTCGAAACTCCACAAACCGGATATCGACAGCGGAAACTTGGGTGCA 316	
Qy	28	ValAlaProValLeuGluLeuThrAlaThrProArgGlnAspAlaAlaGluAlaGly 47	
Db	317	GTCTCCAGCTGCTCTCTCTCTCTCCACCTACAGCAG-----	
Qy	48	ValAspGluProAlaGlnHisGlnCysGluHisPheSerIleArgGlyTyrValAlaLeu 67	
Db	355	-----	-----
Qy	68	LeuGlnLysLysAspProLysPheCysSerLeuSerArgIlePheHisAspGlnLysIys 87	
Db	356	-----AAGCTTCGGCAACTGAAAAAGATCAGAGGAA 388	
Qy	88	CysAspGluHisLysAlaSerSerProPheSerValAlaLysPheArgArgTrpAsp 107	
Db	389	TTGAGCAACTACCCACATCCATTGCGACCGGGTTTCTAATATTACCTCGCCACGT 448	
Qy	108	CysSerLysCysLeuAspLysLeuLysThrSerAspAsnGlyThrAlaProArgThrLeu 127	
Db	449	GTCGCCAGTCGAGCAACGGCTTCAGCAACTAACCCAAATCCAACTTTCCCACAATGCA 508	
Qy	128	ProAlaLys---GlnAsnGlyThrSerAspGlyCysSerIleThrPheValArgSerThr 146	
Db	509	ACATCCAGGCTTCAGACTCCACAGTCAAGATATCGAAATTTGATTTCATCAAGATTGGT 568	
Qy	147	PheValProAlaSerValGlySerGlnLysValSerProSerThrGlnSerSerGlnGly 166	
Db	569	ATCAAGCCAAAGAGCTGTGGACTTTAAACACCCCTCATCATCAACCACCTTCATCAATAAT 628	
Qy	167	LysAsnAlaAspArgSerThrLeuProLysSerValGlnGluGlyAsnAsp----- 183	
Db	629	ACAANTCATTCGGT-----CCGTGAGCGGTTCCAGTGGCATATAAATGTTGGC 679	
Qy	184	-----SerLysCysAsnAlaProSerGlyLysAsnGlyValaAlaGluAla 198	
Db	680	TCGACGATATCCATCTCGCAAGAGCTTAGAATCATCATCAAGCTACAGCTCTATTTCG 739	
Qy	199	AsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAlaAla 218	
Db	740	ATCTAAACCGACTTACCTCCAACTCCAAAACTTCTTACGACCAAAACCAGCTAGTT 799	
Qy	219	AsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIle 238	
Db	800	CGTGTTCGT-----ACAACACAAAAATCGGAAC----- 829	
Qy	239	ThrTrpHisIleGluValAsnGlyValaAspGlnProProSerThrProLysLeuSerGlu 258	
Db	830	-----TCAAGACTAGCCGCTCCGAAAGCCGTGACGACCCCCCAAAATTTGCTTCT 877	
Qy	259	ValVal-----LeuLysArgAsnGluAspGluAsnGlyLysThrGlnGluThrLeu 275	
Db	878	GTGAAGACTATTGAGAGCAAAACAAGAGCCGATACAGCGCTGGTGGTGGTGGATG 937	
Qy	276	ValAlaGluGlnCysAsnLeuThrLysAspPro-----AsnProMetSer 290	
Db	938	CTGAATATTAAAGTTATTTCAGTAGCAAAACCCATCTCTCTCATCGAATAGCCCCAACCT 997	
Qy	291	GlyLysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLysProVal 310	
Db	998	ACGAGAAGCGCGCGGTGCCTCAACACAAACTTTGTGCAAAATCGCTGCCCCAGTG 1057	
Qy	311	---SerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluValValLeuLys 329	
Db	1058	AAAAGTGGCTGAG-----CGCCGACCAAGTAAGCTGGGAAGT 1096	





QY 48 ValAspGluProAlaGlnHisGlnCysGluHisPheSerIleArgGlyTyrValAlaLeu 67  
 Db 346 ----- 346  
 QY 68 LeuGlnLysLysAspProLysPheCysSerLeuSerArgIlePheHisAspGlnLysLys 87  
 Db 347 ----- 347  
 QY 88 CysAspGluHisLysAlaSerSerProPheSerValAlaLysPheArgA-gTTPasp 107  
 Db 380 TTGGAGCAACTACCCACATCATTTATGCCACCGCGGTTCTATAATTTACCTCGCCACGT 439  
 QY 108 CysSerLysCysLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLys 127  
 Db 440 GTGCCACGTCAGCAACCGCTTCAGCAACTAAACCCCTCATCAACCTTCCACAAATGTCA 499  
 QY 128 ProAlaLys---GlnAsnGlyThrSerAspGlyCysSerIleThrPheValArgSerThr 146  
 Db 500 ACATCCAGGCTTCAGACTCCACAGTCAAGATATCGAATATTCGATTCATCAAGATTGGT 559  
 QY 147 PheValProAlaSerValGlySerGlnLysValSerProSerThrThrGlnSerSerGlnGly 166  
 Db 560 ATCAAGCCAAAGAGCTGTGACTTAAACCAACCCCTCATCAACCTTTCATCAATAAT 619  
 QY 167 LysAsnAlaAspArgSerThrLeuProLysSerValGlnGlnGlyAsnAsp----- 183  
 Db 620 ACAANTTCATCCGT-----CCGTGAGCGGTTGAGTGGCAATATAATGTGGC 670  
 QY 184 -----SerLysCysAsnAlaProSerGlyLysAsnGlyAlaAlaGluAla 198  
 Db 671 TCGACGATATCCACATCTCGAAGAGCTTAGAATCATCATCAAGCTACAGCTCTATTTCG 730  
 QY 199 AsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAlaAla 218  
 Db 731 APTCTAAACCGACTACTCTCCAACTCCAAAAACCTTTAGACCCACAAACCCAGCTAGTT 790  
 QY 219 AsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIle 238  
 Db 791 CGTGTTCCT-----ACAACTACAAAAATCGAAGC----- 820  
 QY 239 ThrTrpHisIleGluValAsnGlyAlaAspGlnProProSerThrProLysLeuSerGlu 258  
 Db 821 -----TCAAAGCTAGCGCTCCGAAAGCCGTGAGCACCCCAAACTTGTCT 868  
 QY 259 ValVal-----LeuLysAspAsnGluAspGluAsnGlyLysThrGluGluThrLeu 275  
 Db 869 GTGAAGACTATTGGAGCAAAACAGCCCGATACAGCGGTGTGTGTGTGTGTGTGTGTGTGT 928  
 QY 276 ValAlaGluGlnCysAsnLeuThrLysAspPro-----AsnProMetSer 290  
 Db 929 CTGAATTAAGATTATTTCAGTAGCAAAACCCATCTTCCTCATCGAATAGCCCAACCT 988  
 QY 291 GlyLysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLysProVal 310  
 Db 989 ACGAAGAAAGCGCGCGGCTCTCAACAAACCTTTGCGAAATCGCTGCCCGCCAGTG 1048  
 QY 311 ---SerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluValValLeuLys 329  
 Db 1049 AAAAGTGGCTGAAG-----CGCCGACCAAGTAAAGTGGCAAGT 1087  
 QY 330 ArgSerSerLysSerLys-----ArgLysThrAspLysLys 341  
 Db 1088 GCACGCTGATGTCGAAGCTTTTACGCCAAAAAGTTTCTTACCGTAAACCGGACGCCCA 1147  
 QY 342 LeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAspAla 361  
 Db 1148 ATCATATCTCAACA-----GACTCG 1168  
 QY 362 LysLeuCys-ArgArgLysProLysLys-ValArgLeu-----LeuSerGluIleAla 379  
 Db 1169 AAACGATGCTCAAAGAGCAGTGAAGAGAGTCCGGATACGCTGGATTCAACAGCAAGTC- 1227

QY 379 snAlaAsnGlnValGluAspSerArg-SerAspGluValHisArgGluAsnAlaAsp 398  
 Db 1228 --GCCACGCTCATCGAGCAAGAGTTCCTAAGCATGCAATTCACATCTTCCAAAGAGT 1285  
 QY 399 ProCysGluAspAspArgSerThrIleProValProMetGluValSerMetAspIle--- 417  
 Db 1286 TCAAGCTCAGCAAAAGTCTCCGTCATCAGAGATCTTACTCTTAACGCCCTCCATCGTG 1345  
 QY 418 -----ProValSerAsnHisThrValGlyGluAsp 427  
 Db 1346 ACAGCTATCAGACAGCGGATACCGCAACACCGGTTTCTCCAAATATTATCAACAGCCT 1405  
 QY 428 -----GlyLeuLysSerSerLysAsnLysThrLysArgLysTyr 440  
 Db 1406 GTTGAGGAAAAACCAACACTGCGCATGGAAGAGTGAAGCAACAGCAGCAAAAGATCCA 1465  
 QY 441 SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysLysArg 460  
 Db 1466 CCTCCAGCTGTTCCGCCAGTGCACAC-----CAGCCAACA 1501  
 QY 461 ThrGlySerValHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
 Db 1502 ATCGGAGTTTGTAGTCCAATTATGACACAT-----AAGAAGTTG 1540  
 QY 481 Thr-----ProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeu 498  
 Db 1541 ACAATGACCCCGTGTATCTGAAAAACCAAGAACCTGAAAAGCTCCAATCAATGAGCATC 1600  
 QY 499 AspThrAsnMetHisLysThrAspVal---CysGlnHisValSerGluLysSerThrGln- 517  
 Db 1601 GACAGC-----ACGACGTTCCACCGCTTCCACCTCTAAATCACTTGTTCACA 1648  
 QY 518 -----ArgCysSerSerLysGlyLysThrA 526  
 Db 1649 CTTAAAAATGACTTCAATCCGACACACCAACGCTAGATGTTCTTCTAAACAGG----- 1704  
 QY 526 laGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysTyrGlyGlyLysSer 546  
 Db 1705 -----AAAAATCACATCGCTGTCAAG--TCGTTTGGATATGAGCAT 1745  
 QY 546 hArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrG 566  
 Db 1746 CG-----TCCGCGCTGT 1757  
 QY 566 luAsnSerValLeuSerHis---SerAlaLysValSerPro-----AlaG 580  
 Db 1758 AAGACTCCATTGTGCTCATGCTCGGCTCAGTCTCGCGCCGACAAAAAATCTTGTA 1817  
 QY 580 luHisAspIleGlnIleMetSerAspLeuHisGlnSerLeuProLysLysLysLysL 600  
 Db 1818 ATCATTCGCTG-----GAGAGAAGGATGGAAGAATAAGACAT 1856  
 QY 600 ysGlnLysLeuGluValThr-----AspGluL 609  
 Db 1857 CAGATCCAGCGGCTACACTCTGACGCGGTTGGTGGATGCGCCCAAAATGAGGAGA 1916  
 QY 609 ysGlnThrMetIleAspIleProMetAspIleValIleuLeuAlaLysAsnGlnH 629  
 Db 1917 AGCTCAAGAAATACGATGACATGACTCT-----CGAGCACAGACGCT 1961  
 QY 629 is-----GluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgI 645  
 Db 1962 ATCTCGAACACTTCGAAGACAGTTCCTCTGTCTGGATATCCGATATCCGATACAGAGC 2021  
 QY 645 laGlnSerLysThrThrAlaAsp-----AspAspCysValIleValAlaAlaLysA 662  
 Db 2022 TCGACGACATATCCAGCGACGATTTGTCGGAGTAGACATGCAACAGTGCCTCCAAAC 2081  
 QY 662 spGlySerAspTyrAlaSerSerVal---PheAspThrAsnSerGlnGlnLys----- 678  
 Db 2082 AT--AGGAGCTATTCCCACTTTGTTCGCCATCCACGCTCTTCTTCTCCAAAGCCCGAG 2138  
 QY 679 -----SerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrT 697

Db 2139 TCCCAAGTCGTCCTCCATCATGATCGATTCGATCTCGACGAGAA----- 2185  
Qy 697 hGlnGlnSerProHisProGlnAsnPhelGlnSerThrGlnGlnGlnGlnHisLeuA 717  
Db 2186 --CAGGAGAATGTGTACAAACTTCTGTCGACGCGAAGCAGCAGCAACGTGGC----- 2236  
Qy 717 rGmetGluGluMetValThrIleAlaAlaSerSerProLeuPheSerHisHis----- 734  
Db 2237 -----GCGCTGCCACCTCAACCTTCGGACCAACATCGCTAA 2273  
Qy 735 --AspAspGlnTyrIleAlaGluAlaProThrGluHisTyrGlyArgLysAspAlaLysL 754  
Db 2274 GATCCCGCGGATCACTCATCTATCTCCCA-----CACTTATCA---GTGTCAGCTGATA 2324  
Qy 754 yLeuThrTrpGluGlnPheLysAlaThrThrArg-----AsnSerProAlaAlaThrC 772  
Db 2325 AGACACACATGCTATGCTACACAGCTAGTCGACGACCTTCTTCACAAAACCAAGCT 2384  
Qy 772 yGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisValMetG 792  
Db 2385 ATTCAGGCCAAATTT-----CACTCACTTGAT---CGTAAATGCCACCTTCAG 2429  
Qy 792 lySerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAspArgTyrAlaG 812  
Db 2430 AGTTCATCACTCCAGCAGCAGATGCGCGCTCTCTTGAGCCCG----- 2473  
Qy 812 luArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGluA 832  
Db 2474 -----AGACGGGTGCCGAACCTCGATGTCGAAATATGATT 2507  
Qy 832 laSerIysLeuCysAspArgArgAsnAlaGly-----GlnValValLeuTyrProLysG 850  
Db 2508 CTTCAAGATCTTACTCGCGCGCTTCCGAGGTGGAGCTCTACTGTATCTATGAGAG- 2566  
Qy 850 luSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPheP 870  
Db 2567 -----ACGTTCCACTGCACAGACTATCCGATGAAAATCCCGGCACAT---- 2611  
Qy 870 roAsnTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerGlnTyrA 890  
Db 2612 -----TCTGCCAAAGTGAGATGGATCCCACTA----- 2641  
Qy 890 laHisAsnGlnTyrLysGlySerThrSerThrSerThrSerThrSerThrSerThrSer 910  
Db 2642 -----TCAGTGGCTAGCAGCAGCAGCATATGATCT---CTCAATGAGAGT 2684  
Qy 910 leProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgProLeu 929  
Db 2685 AC-----GAACATGCTATTCCGACATGCGACATGCGACCTGACTTGG 2720  
Qy 930 -----ArgProHisProArgValGlyValLeuG 939  
Db 2721 AGTGTACAGAACACTGTCGACTCTAACCAGAAACAGAGAGAACTATGAGCAATTGT 2780  
Qy 939 lySerLeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSer---- 957  
Db 2781 TTGATCTTTTGAGCAAAAGCTTAGAAAACCTCACTCAACACATTTGATCGATCCAACTGA 2840  
Qy 958 -----GlyTyrLysLeuGlyValSerThrG 966  
Db 2841 AGCCTGAAGAGGCAATACGATTTCAGCAGCAGCATTTGCTCATTTGAGGGATATTAGCAATC 2900  
Qy 966 lyIleThrSerHisGlnMetAsnArgLysGlu----- 976  
Db 2901 ATCTTCATCCAACTCAGCTCATGCTAAACGAGGCGCTGCTGAGCTTCTTCGTCAACCAT 2960  
Qy 977 -----HisPheGluAlaLeuAsnSerGlyMetPheSerAlaLysT 990  
Db 2961 CTTCTGAATCAGTTGTCATCCCATCATCATCATGTCATGTCGTCGTCGTCGTCGTCGTCG 3020  
Qy 990 rpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAspPheLeuSerAlaArgAsnS 1010

Db 3021 AGGAGAAGATCAGCTTGCAGCTGCTTGGCAAGAAC----- 3055  
Qy 1010 erIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHisProLeuAspArgPheV 1030  
Db 3056 -----AAGAAGAGCTGGATCGCTCCTCA-----CTCTCCAGTTCA 3092  
Qy 1030 alArgGlnAspIleCysIleThrAsnLysAsn-----ProAlaAspPheThrThrIleS 1048  
Db 3093 CCAAGAAG-----AAGAACAAGAACTACGACGAGCAGCATATGCTCATCAATT 3140  
Qy 1048 erAsnAspAsnGluTyrMetAsp 1055  
Db 3141 CCGATCTCAAGGAACCTCTTGAC 3163  
RESULT 12  
US-09-845-917A-26  
; Sequence 26, Application US/09845917A  
; Patent No. 6653529  
; GENERAL INFORMATION:  
; APPLICANT: Bogaert, Thierry  
; APPLICANT: Vandekerckhove, Joel  
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH  
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND  
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR  
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR  
; FILE REFERENCE: P/14-1  
; CURRENT APPLICATION NUMBER: US/09/845,917A  
; CURRENT FILING DATE: 2001-04-30  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 7474  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: artificial  
; OTHER INFORMATION: plasmid  
US-09-845-917A-26  
Alignment Scores:  
Pred. No.: 2,466-09 Length: 7474  
Score: 211.00 Matches: 254  
Percent Similarity: 34.74% Conservative: 174  
Best Local Similarity: 20.62% Mismatches: 415  
Query Match: 3.82% Indels: 390  
DB: 4 Gaps: 62  
US-09-721-114-2 (1-1057) x US-09-845-917A-26 (1-7474)  
Qy 8 GlnGluGlyAlaArgValGlyThrAsnCysMetLeuAlaArgGlyGlyThrGlyAla 27  
Db 347 AAGGAAGGAGAAAGCGAAGAGAGCGGC-----GCTAGG---GCGCTGCCAAGT 394  
Qy 28 ValAlaProValLeuGluLeuThrAlaThrProArgGlnAspAlaAlaGluAlaGly 47  
Db 395 GTACGGTTCAGCTGCGCGTAACCAACCAACACCC-----GCCGCG 433  
Qy 48 ValAspGluProAlaGln-----HisGlnCys---Glu 57  
Db 434 CTTAATCGCGCTACAGGCGCGTCCCATTCAGCTTCGCACTGTGGCAA 493  
Qy 58 HisPheSerIleArgGlyTyrValAlaLeuLeuGlnLysAspProLysPheCysSer 77  
Db 494 GGGCGATCGTGGCGGCC---TCTTCGCTATTACGCCAGCTGGCGAAGGGGGATGTGCT 550  
Qy 78 LeuSerArgIle-----PheHisAspGlnLysLysCysAspGluHis 91  
Db 551 GCAAGGCGATTAACTGGGTAAACCCAGGGTTTCCCGAGTCACGACGTTGTAACACGACG 610  
Qy 92 LysAlaSerSerSerProPheSerValAlaLysPheArgTrpAspCysSerLysCys 111  
Db 611 GCCAGTGAAGCGCGGTAATAC----- 631



Db 2487 -----GCCGCTGCCACCTCAACCTCCGACCAACATTCGCTA 2522  
QY 735 ---AspAspGlnTyrIleAlaGluAlaProThrGluHisTrpGlyArgLysAspAlaLys 753  
Db 2523 AGATCCCGGGATACCTATCTCTCA-----CACTTATCA-----GTGTCACTGAT 2573  
QY 754 LysLeuThrTrpGluGlnPheLysAlaThrThrArg-----AsnSerProAlaAlaThr 771  
Db 2574 AAGGACAACTGCTATGCACTACAGACTAGTCAGACCTCTCTCAAAAACCAAGC 2633  
QY 772 CysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisValMet 791  
Db 2634 TATTCAGGCGCAATTT-----CACTCACTGAT---CGTAATGCCACCTTCAA 2678  
QY 792 GlySerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAspArgTyrAla 811  
Db 2679 GAGTTCACATCCAGCAGACAGATGGGGCTCTCTTGAGCCGG----- 2723  
QY 812 GluArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGlu 831  
Db 2724 -----AGACGGGTGCCGAACTCGATGTGCGAAATATGAT 2756  
QY 832 AlaSerLysLeuCysAspArgAsnAlaGly-----GlnValValLeuTyrProLys 849  
Db 2757 TCTTCAGGATCTACTCGCGGGTCCCGAGTGGAGCTCTACTGGTATCTATGGAGAG 2816  
QY 850 GluSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSerPhe 869  
Db 2817 -----ACGTTCCAACTGCACAGACTATCGGATGAAATAATCCCGCCACAT--- 2861  
QY 870 ProAsnTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerGlnTyr 889  
Db 2862 -----TCTCCAAAGTGAGATGGGATCCCACTA----- 2891  
QY 890 AlaHisAsnGlnTyrLysGlySerThrSerThrSerThrTyrGlySerAsnLeuAsnGlyLys 909  
Db 2892 -----TCACTGGCTAGCAGCAGCATATGATCTCT---CTCAATCAGAG 2933  
QY 910 IleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgProLeu 929  
Db 2934 TAC-----GAACATGCTATTCCGGACATCCGACGTCACCTTG 2969  
QY 930 -----ArgProHisProArgValGlyValLeu 938  
Db 2970 GAGTGTACAGAACACTGCTGACTCACTAACCAAGAAACAGAGAACTATGAGCATTTG 3029  
QY 939 GlySerLeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSer--- 957  
Db 3030 TTTGATCTTTTTCAGCAAAAGCTTAGAAAACTCACTCAACACATTTGATCGATCCAACTTG 3089  
QY 958 -----GlyTyrLysLeuGlyValSerThr 965  
Db 3090 AAGCCTGAGAGGCATATACATTACAGCAGGACATTTGCTCATTTGAGGGATATTAGCAAT 3149  
QY 966 GlyIleThrSerHisGlnMetAsnArgLysGlu----- 976  
Db 3150 CATCTTGCATCCAACTCAGCTCATGCTAACGAGGCGCTGGTGGAGCTTCTTCTGTCACCA 3209  
QY 977 -----HisPheGluAlaLeuAsnSerGlyMetPheSerAlaLys 989  
Db 3210 TCTCTGGAATCAGTTGATCCCATCCATCATTCATGCTATGCTGTCGAAAGCAGCAAG 3269  
QY 990 TrpAsnAlaLeuGlnLeuGlySerValSerSerSerAlaAspPheLeuSerAlaArgAsn 1009  
Db 3270 CAGGAGAAGATCAGCTTGAGCTGCTTGGCAGAAC----- 3305  
QY 1010 SerIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHisProLeuAspArgPhe 1029  
Db 3306 -----AAGAAGAGCTGGATCGCTCTCTCA-----CTCTCAAGTTC 3341  
QY 1030 ValArgGlnAspIleCysIleThrAsnLysAsn-----ProAlaAspPheThrThrIle 1047  
Db 3342 ACCAAGAG-----AAGAACAAGAACTACGACGAGAACACATATATGCCATCAAT 3389

QY 1048 SerAsnAspAsnGlnTyrMetAsp 1055  
Db 3390 TCCGATCTCAAGCACTCTTGAC 3413

## RESULT 13

US-09-845-917A-49  
; Sequence 49, Application US/09845917A  
; Patent No. 6653529  
; GENERAL INFORMATION:  
; APPLICANT: Bogaert, Thierry  
; APPLICANT: Vandekerckhove, Joel  
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH  
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND  
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR  
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR  
; FILE REFERENCE: P/14-1  
; CURRENT APPLICATION NUMBER: US/09/845,917A  
; CURRENT FILING DATE: 2001-04-30  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 49  
; LENGTH: 4217  
; TYPE: DNA  
; ORGANISM: Caenorhabditis elegans  
US-09-845-917A-49  
Alignment Scores:  
Pred. No.: 3,51e-09 Length: 4217  
Score: 205.00 Matches: 232  
Percent Similarity: 34.86% Conservative: 156  
Best Local Similarity: 20.84% Mismatches: 379  
Query Match: 3.71% Indels: 347  
DB: 4 Gaps: 54  
US-09-721-114-2 (1-1057) x US-09-845-917A-49 (1-4217)

QY 110 LysCysLeuAspLysLeuLysThrSerAspAsnGlyThrAlaProA-gThrLeuProAla 129  
Db 6 AAGAAATGGAGCAACTACCCACATCCATTATGCGCCGGGTTTCTAAATTACCTCG 65  
QY 130 LysGlnAsnGlyThrSerAspGlyCysSerIleThrPheValArgSerThrPhe--- 147  
Db 66 CCAGTGTGCGCAGCTCAGCAACCGCTTCAGCAACTAACCCAAATTCCACTTTCCCAA 125  
QY 147 ----- 147  
Db 126 ATGTCAACATCCAGCTTCAGACTCCACAGTCAAGAAATATCGAAATTTGATTCAATAAG 185  
QY 148 -----ValProAlaSerValGlySerGlnLysValSerProSerThrGlnSerSer 164  
Db 186 ATTGGTATCAAGCCCAAGACGCTGAGACTTAACACCCCTCATCATCAACCACTTCATCA 245  
QY 165 GlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGlnGluGlyAsnAsp--- 183  
Db 246 AATATACAAATTCATTCCTG-----CCGTGAGCGCTTCGAGTGGCAATAATAAT 296  
QY 184 -----SerLysCysAsnAlaProSerGlyLysAsnGlyAlaAla 196  
Db 297 GTTGGCTCGACGATATCCACATCTCGAAGAGCTTAGAATCATCATCAACGTACAGCTCT 356  
QY 197 GluAlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspVal 216  
Db 357 ATTTGGAATCTTAACCGACCTACCTCCCACTCCAAAACCTCTTAGACCAACCAACCCAG 416  
QY 217 AlaAlaAsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValPro 236  
Db 417 CTAGTTCTGTTGCT-----ACAACTCAAAAATCGGAAGC----- 452  
QY 237 GlnIleThrTrpHisIleGluValAsnGlyAlaAspGlnProProSerThrProLysLeu 256  
Db 453 -----TCAAGCTAGCCCTCGAAGAGCGGTGAGCCGACCCCAAACTT 494

Qy	257	SerGluValVal	-----LeuLysArgAsnGluAspGluAsnGlyLysThrGluGlu	273
Db	495	GCTTCTCTGAAGACTATTGGAGCAAAACAAGAGCCGATACACAGCGTGGTGGTGGTGGT	554	
Qy	274	ThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro	288	
Db	555	GGAAAGTGTGAATTAAGATTATTACAGTAGCAAAACCCATCTTCTCTCATCGAATAGCCCA	614	
Qy	289	MetSerGlyLysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLys	308	
Db	615	CAACCTACGAGAAAGGGGGCGGTGCTTCAACACAAACTTTGTTCGAAAAATCGTGTCC	674	
Qy	309	ProVal---SerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluValVal	327	
Db	675	CCAGTGAAGATGCGCTGAAG-----CCGCGACCACTGACGCTG	713	
Qy	328	LeuLysArgSerSerLysSerLys-----ArgLysThrAsp	339	
Db	714	GGAAAGTCCACGCTCTATGTCGAAGCTTTGTAGCGCAAAAGTTTCTACCGTAAACCGGAC	773	
Qy	340	LysLysLeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSer	359	
Db	774	GCOCCTAATCATATCTCAACAA-----	794	
Qy	360	AspAlaLysLeuCys-ArgArgLysProLysLys-ValArgLeu-----LeuSerGluI	377	
Db	795	GACTCGAAACGATGCTCAAGAGCAGTCAAGAGAGTCCGATACGCTGGATTCATCAACAGC	854	
Qy	377	IleLeuAlaAsnGlnValGluAspSerArg-SerAspGluValHisArgGluAsnAla	396	
Db	855	ACGTG---GCCAACGTCTCATCGACGAAGTTCCCTAAGCAATGCATTCACATCTTCC	911	
Qy	397	AlaAspProCysGluAspAspArgSerThrIleProValProMetGluValSerMetAsp	416	
Db	912	AAGAGTTCACGTCAGACGAAAGTCTCCGTTCATCAGACGATCTTACTCTTAACGCTCC	971	
Qy	417	Ile-----ProValSerAsnHisThrValGly	425	
Db	972	ATCGTGACAGTATCAGACACCGCATAGCCGCAACACCGGTTTCTCCAAATATTATCAAC	1031	
Qy	426	GluAsp-----GlyLeuLysSerSerLysAsnLysThrLysArg	438	
Db	1032	AAGCCTTTGAGGAAACCAACCACTGCGAGTGAAGAGAGTGAAGACAGACAGACGAAAAAA	1091	
Qy	439	LysTyrSerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLys	458	
Db	1092	GATCCACCTCCAGCTGTTCCGCCAGTGACACC-----CAG	1127	
Qy	459	LysArgThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLys	478	
Db	1128	CCAACAATCGAGTGTGTAGTCCCAATTATGCGACAT-----AAG	1166	
Qy	479	LysValThr-----ProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsn	496	
Db	1167	AAGTTGACAAATGACCCCGTGATATCTGAAACACAGAACCTGAAGAGTCCCAATCAATG	1226	
Qy	497	GlyLeuAspThrAsnMetHisLysThrAspVal---CysGlnHisValSerGluIleSerT	516	
Db	1227	AGCATGCACACG-----ACGAGAGTTCACCGCTTCACCGCTTCCAAATCAAGTT	1274	
Qy	516	hrGln-----ArgCysSerSerLysGlyL	524	
Db	1275	GTTCCACTTAAATGACTCTCAATCCGACACCAACCACTGACGATGTTCTCTTAAACAA	1334	
Qy	524	ysThrAlaGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysTyrGlyGlyG	544	
Db	1335	GG-----AAAAATCACATCGCGCTGTCAAG---TCGTTTGGATATG	1371	
Qy	544	luSerThrArgAsnGlyAlaAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetG	564	
Db	1372	AGCAGTGG-----TCOG	1383	
Qy	564	luThrGluAsnSerValLeuSerHis---SerAlaLysValSerPro-----	578	

Db	1384	CGTCTGAAGACTCCATTGTGGCTCATGTCGGCTCAGGTGATCTCCGCCGACAAAACCTT	1443
Qy	579	--AlaGluHisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLeuLysL	598
Db	1444	CTGTAATCATTCGCTG-----GAGAGAAGCATGGGAAGAATA	1482
Qy	598	ysylsLysGlnLysLeuGluValThr-----A	607
Db	1483	AGACATCAGAATCCAGCGGTACATCTGACCGCGGTGTGGCATGTGCGCCAAATGA	1542
Qy	607	rgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysA	627
Db	1543	GGGAGAAGCTGAAGAATACGATGACTCGT-----CGAGCACAGA	1587
Qy	627	snGlnHis-----GluArgGlnLeuMetThrGluThrAspCysSerAspIleA	643
Db	1588	ACGGCTATCTCGAACAATTCGAAGACAGTTTCCTCCTTGTGCTCGGAATATCCGATAACA	1647
Qy	643	snArgIleGlnSerLysThrAlaasp-----AspAspCysValIleValAlaA	660
Db	1648	ACGAGCTCGACACATATCCACGACGATTTGTCGGAGTAGACATGGCAACATGCGCT	1707
Qy	660	lalyAspGlySerAspTyrAlaSerVal---PheAspThrAsnSerGlnGlnLys-	678
Db	1708	CCAAACAT---AGCGCATATTCCCATTTGTCGCCATCCACGTCTTCTTCCTCAAAGC	1764
Qy	579	-----SerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaL	695
Db	1765	CCGGAGTCCCGAGTCGGTCTCCACATCAGTCGATCTCGATCTCGAGCAGAA-----	1817
Qy	695	euThrThrGlnGluSerProHisProGlnAsnPheGlnSerThrGlnGluGlnGlnThrH	715
Db	1818	-----CAGGAGAATGTACAAACTTCGTGCCAGTCCGACGCGCAACGACGTCGGC-	1868
Qy	715	isleuArgMetGluGluMetValThrIleAlaAlaSerSerProLeuPheSerHisHis-	734
Db	1869	-----GCCCTGCCACCTCAACTTCGAGCACATTT	1899
Qy	735	-----AspAspGlnTyrIleAlaGluAlaProThrGluHisTrrPglyArgLysAspA	752
Db	1900	CGCTAAGATCCCGGGATCATCTATCTTCCA-----CACTTATCA-----GTGTCAG	1950
Qy	752	lalyLysLysLeuThrTrpGluGlnPheLysAlaThrThrArg-----AsnSerProAlaA	770
Db	1951	CTGATTAAGGACACAATGTCTATCACTATCAGACTAGTCGACGACCTTCTTCACAAAAAC	2010
Qy	770	laThrCysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisV	790
Db	2011	CAGCTATTACGCCAATTT-----CATTCACCTGAT-----CGTAATGCCACC	2055
Qy	790	alMetGlySerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAspArgT	810
Db	2056	TTCAAGAGTTTCACATCCACCGACACAGATGCGGCTCTCTTGAGCCCG-----	2105
Qy	810	yrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrM	830
Db	2106	-----AGCGGGTCCGAACTCGATGCCAAT	2133
Qy	830	etGluAlaSerLysLeuCysAspArgArgAsnAlaGly-----GlnValValLeuTyrP	848
Db	2134	ATGATTCTTCAGGATCTACTCGCGCGTTCGCCAGGTGGAAGCTCTACTGGTATCTATG	2193
Qy	848	roLysGluSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlas	868
Db	2194	GAGAG-----ACGTCCCACTCCACAGACTATCCGATGAAAAATCCCCCGCAC	2241
Qy	868	erPheProAsnTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerG	888
Db	2242	AT-----TCGTCCAAAAGTGAGATGGGATCCCAACTA-----	2273
Qy	888	lnTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerThrGlySerAsnLeuAsnG	908

Db 2274 -----TCACTGGCTAGCAGCAGCATATGATGCT---CTCAATG 2310  
Qy 908 lylsyleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgp 928  
Db 2311 AGAAGTAC-----GAACATGCTATCGGACATCGGCAGCTG 2346  
Qy 928 rleu-----ArgProHisProArgValGlyV 937  
Db 2347 ACTTGAGGTGTACAGAACACGTGTCGACTCACTAACCAAGAACAGGAGAACTATGGAG 2406  
Qy 937 alileGlySerLeuLeuGlnLysGluAlaAsnTrpSerGluAnCysGlyThrGlns 957  
Db 2407 CATGTTGATCTTTTGGACAAAGCTTAGAAACTCACTCAACACATGATGATCCA 2466  
Qy 957 er-----GlyTyLysLeuGlyVals 964  
Db 2467 ACTTGAGCCTGAGAGGCAATACGATCAGCCAGCAGCATTCCTCATTTGAGGGATATTA 2526  
Qy 964 erThrGlyThrSerHisGlnMetAspArgLysGlu----- 976  
Db 2527 GCAATCATCTTGATCACTCACTCAGCTCATGCTAACGAGGGCGTGTGAGCTTCCTGTC 2586  
Qy 977 -----HisPheGluAlaLeuAnsnSerGlyMetPheSerA 988  
Db 2587 AACCATCTCTGGAATCAGTTGCATCCCATGATCATCGATGTCATGCTGTCGAAAGCA 2646  
Qy 988 lalysTrpAsnAlaLeuGlnLysSerValSerSerSerAlaAspPheLeuSerAlaA 1008  
Db 2647 GCAGCAGGAGAGATCAGCTTCAGCTGTTTGGCAGAAC----- 2687  
Qy 1008 rGAsnSerileAlaGlnSerThrArgGlyLysGlyLysMetValHisProLeuAspA 1028  
Db 2688 -----AAGAAGAGCTGATCGCTCTCA-----CTCTCCA 2718  
Qy 1028 rGpHeValArgGlnAspIleCysIleThrAsnLysAsn-----ProAlaAspPheThrT 1046  
Db 2719 AGTTCACCAAGAG-----AAGAACAGAACTACGACGAGCAGCATATGCCAT 2766  
Qy 1046 hrileSerAsnAspAsnGluTyMetAsp 1055  
Db 2767 CAATTCGGATCTCAAGGAACCTTGAC 2795

RESULT 14  
US-09-845-917A-53  
; Sequence 53, Application US/09845917A  
; Patent No. 6653529  
; GENERAL INFORMATION:  
; APPLICANT: Bogaert, Thierry  
; APPLICANT: Vandekerckhove, Joel  
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH  
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND  
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR  
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR  
; FILE REFERENCE: P/14-1  
; CURRENT APPLICATION NUMBER: US/09/845,917A  
; CURRENT FILING DATE: 2001-04-30  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 53  
; LENGTH: 4382  
; TYPE: DNA  
; ORGANISM: Caenorhabditis elegans  
US-09-845-917A-53

Alignment Scores:  
Pred. No.: 3,75e-09 Length: 4382  
Score: 205.00 Matches: 232  
Percent Similarity: 34.86% Conservative: 156  
Best Local Similarity: 20.84% Mismatches: 379  
Query Match: 3.71% Indels: 347  
DB: 4 Gaps: 54

US-09-721-114-2 (1-1057) x US-09-845-917A-53 (1-4382)

Qy 110 LysCysLeuAspLysLeuLysThrSerAspAsnGlyThrAlaProArgThrLeuProAla 129  
Db 6 AAGAAATTTGGAGCAACTACCCACATCCATATGACCCCGGCTTTCTAAATATACCTTCG 65  
Qy 130 LysGlnAsnGlyThrSerAspGlyCysSerIleThrPheValArgSerThrPhe 147  
Db 66 CCAGTGTGGCAGCTCAGCAACCGCTTCAGCAACTAACCCAAATTCCAACTTCCACA 125  
Qy 147 ----- 147  
Db 126 ATGTCACATCCAGCTTCAGACTCCACAGTCCAGAAATATCGAAAATGATTCATCAAG 185  
Qy 148 -----ValProAlaSerValGlySerGlnLysValSerProSerThrGlnSerSer 164  
Db 186 ATTGTATCAAGCCCAAGACGCTGACTTAAACCCACCTCATCATCAACCACTTCATCA 245  
Qy 165 GlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGlnGluGlyAsnAsp 183  
Db 246 AATAATACAAATTCATTCGGT-----CCGTGAGCCGCTCGAGTGGCAATTAAT 296  
Qy 184 -----SerLysCysAsnAlaProSerGlyLysAsnGlyAlaAla 196  
Db 297 GTGGCTCGAGATATCCACATCTCGAGAGAGCTTAGAATCATCATCAAGTACAGTCT 356  
Qy 197 GluAlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspVal 216  
Db 357 ATTTCCGAATCTAAACCCGACCTACCTCCCAACTCCAAAACCTTCTAGACCAACACCCAG 416  
Qy 217 AlaAlaAsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValPro 236  
Db 417 CTAGTTCGTGTGCT-----ACAATCAAAAATTCGGAGC----- 452  
Qy 237 GlnIleThrTrpHisIleGluValAsnGlyAlaAspGlnProProSerThrProLysLeu 256  
Db 453 -----TCAAGCTAGCCGCTCGAAAGCGGTGAGCAGCCCAAACTT 494  
Qy 257 SerGluValVal-----LeuLysArgAsnGluAspGluAsnGlyLysThrGluGlu 273  
Db 495 GCTTCTGTGAAGACTATTGGAGCAAAACAGAGCCCGGATACAGCGGTGGTGGTGGT 554  
Qy 274 ThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro-----AsnPro 288  
Db 555 GGAATGCTGAATTAAGTATTTCAGTAGCAAAAACCCATCTTCCTCATCGAATAGCCCA 614  
Qy 289 MetSerGlyLysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLys 308  
Db 615 CAACCTACGAGAAAGCGCGCGGTGCTCAACCAACAACTTTGTCGAAAATTCGCTGCC 674  
Qy 309 ProVal-----SerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluValVal 327  
Db 675 CCAGTGAAGAGTGGCTGAG-----CCGCGCAGCAGTAAGCTG 713  
Qy 328 LeuLysArgSerSerLysSerLys-----ArgLysThrAsp 339  
Db 714 GGAAGTGGCCAGCTGCTATGTCGAAGCTTTGTACGCCAAAAGTTTCTACCGTAAAAACGGAC 773  
Qy 340 LysLysLeuMetLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSer 359  
Db 774 GCCCAATCATATCTCAACAA----- 794  
Qy 360 AspAlaLysLeuCys-ArgArgLysProLysLys-ValArgLeu-----LeuSerGluI 377  
Db 795 GACTCGAAGACGATGCTCAAGAGCAGTGAAGAGAGTCCGATACGCTGGATTCAACAGC 854  
Qy 377 leileAsnAlaAsnGlnValGluAspSerArg-SerAspGluValHisArgGluAsnAla 396  
Db 855 AGCTC-----GCCAAGCTCATCATCGACGAGGAGTCCCTTAAGCATGATTCACATCTCC 911  
Qy 397 AlaAspProCysGluAspAspArgSerThrIleProValProMetGluValSerMetAsp 416  
Db 912 AAGAGTTCAACGCTCAGACGAGAAAGTCTCGTTCATCAGACGATCTTACTCTTAACGCTCC 971





```

Db 2719 AGTTCACCAAGAAG-----AAGAACAAAGAACTACGACGAAAGCACATATGCCAT 2766
Qy 1046 hrileSerAsnAspAsnGluTyrMetAsp 1055
Db 2767 CAATTCGGGATCTCAAGAACTCTTGAC 2795

RESULT 15
US-09-845-917A-50
; Sequence 50, Application US/09845917A
; Patent No. 6653529
; GENERAL INFORMATION:
; APPLICANT: Bogaert, Thierry
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/845.917A
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 4642
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-845-917A-50

Alignment Scores:
Pred. No.: 4,15e-09 Length: 4642
Score: 205.00 Matches: 232
Percent Similarity: 34.86% Conservative: 156
Best Local Similarity: 20.84% Mismatches: 379
Query Match: 3.71% Indels: 347
DB: 4 Gaps: 54

US-09-721-114-2 (1-1057) x US-09-845-917A-50 (1-4642)
Qy 110 LysCysLeuAspLysLeuLysThrSerAspAsnGlyThrAlaProArgThrLeuProAla 129
Db 6 AAGAAATGGAGCACTACCCACATCCATTATGACCACCCCGGTTCTCAANTTACCCTCG 65
Qy 130 LysGlnAsnGlyThrSerAspGlyCysSerIleThrPheValArgSerThrPhe 147
Db 66 CCAGCTGCGCAGCTCAGCAACCGCTTCAGCAACTAACCCAAATTCACACTTTCACAA 125
Qy 147 ----- 147
Db 126 ATGTCAACATCCAGGCTTCAGACTCCACAGTCAAGATATCGAAATTTGATTCATCAAG 185
Qy 148 -----ValProAlaSerValGlySerGlnLysValSerProSerThrGlnSerSer 164
Db 186 ATGGTATCAAGCCAAAGACGCTGGACTTAAACCAACCCCTCATCATCAACCACTTCATCA 245
Qy 165 GlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGlnGluGlyAsnAsp 193
Db 246 AATAATAACAATTCATTCCTG-----CCGTGAGCGGTTTCGAGTGGCAATAATAAT 296
Qy 184 -----SerLysCysAsnAlaProSerGlyLysAsnGlyAlaAla 196
Db 297 GTTGCTCGAGTATCCACATCTCGAAGAGCTTAGAATCATCATCAAGCTACAGCTCT 356
Qy 197 GluAlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnThrAspVal 216
Db 357 ATTTGGAATCTAAACCGACCTTACCTCCAACTCCAAACACCTTCTAGACCACCAACCCAG 416
Qy 217 AlaAlaAsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValPro 236
Db 417 CTAGTTCGTGTGCT-----ACAACTACAAAATCGAAGC----- 452
Qy 237 GlnIleThrTrpHisIleGluValAsnGlyAlaAspGlnProSerThrProLysLeu 256
Db 453 -----TCAAGAGTAGCGGCTCCGAAAGCCGTCGAGCAGCCCAAACTT 494

```

```

Qy 257 SerGluValVal-----LeuLysArgAsnGluAspGluAsnGlyLysThrGluGlu 273
Db 495 GCTTCTGTGAAGACTATTGGAGCAAAACAGAGCCGCGATAACAGCGTGTGTGTGTGTGT 554
Qy 274 ThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro-----AsnPro 288
Db 555 GGAATGCTGAAATTAAAGTTATTTCAGTAGCAAAAACCCATCTTCTCATCGAATAGCCCA 614
Qy 289 MetSerGlyLysGluAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLys 308
Db 615 CAACTACGAGAAAGCGCGCGCTCTCAACAACAAACTTTGCGAAAATCGCTGCC 674
Qy 309 ProVal---SerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluVal 327
Db 675 CCAAGTCAAAAGTGGCTGAAG-----CCGCGGACCACTAGAGCTG 713
Qy 328 LeuLysArgSerSerLysSerLys-----ArgLysThrAsp 339
Db 714 GGAAGTGCACCGTCTATGTCGAAGTTTGTACGCCAAAGTTTCTTACCGTAAACGGAC 773
Qy 340 LysLysLeuMetLysGlnGlnHisSerLysArgThrAlaGlnAlaAspValSer 359
Db 774 GCCCAATCATATCTCAACA----- 794
Qy 360 AspAlaLysLeuCys-ArgArgLysProLysLys-ValArgLeu-----LeuSerGlu 377
Db 795 GACTCGAAACGATGCTCAAGAGCAGTGGAAGAGTCCGATACGCTGGATTCAACAGC 854
Qy 377 IleLeuAsnAlaAsnGlnValGluAspSerArg-SerAspGluValHisArgGluAsnAla 396
Db 855 AGGTC---GCCAAGCTCATCATCGCGGAGGTTCCTTAAGCATGATTCACATCTTCC 911
Qy 397 AlaAspProCysGluAspAspArgSerThrIleProValProMetGluValSerMetAsp 416
Db 912 AAGAGTTCAACGTCAGACGAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCTCC 971
Qy 417 Ile-----ProValSerAsnHisThrValGly 425
Db 972 ATCGTACAGCTATCAGACGCGGATAGCGCAACACCGGTTTCTCCAAATATTATCAAC 1031
Qy 426 GluAsp-----GlyLeuLysSerSerLysAsnLysThrLysArg 438
Db 1032 AAGCGTGTGTGAGGAAAAAACCAACACTGCGAGTGAAGGAGTGAAGAACACACGCAAAAA 1091
Qy 439 LysTyrSerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLys 458
Db 1092 GATCCACCTCCAGCTGTTCCGCGCAGTGACAC-----CAG 1127
Qy 459 LysArgThrGlySerValHisThrValAlaHisProAlaGlyAsnLeuSerAsnLys 478
Db 1128 CCAACAATCGGAGTTGTTAGTCCAATTATGGCACAT-----AAG 1166
Qy 479 LysValThr-----ProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsn 496
Db 1167 AAGTTGACAAATGACCCCGGTATATCTGAAAACACAGAACCTGAAAAGCTCAATCAATG 1226
Qy 497 GlyLeuAspThrAsnMetHisLysThrAspVal---CysGlnHisValSerGluLysSer 516
Db 1227 AGCATCGACAG-----ACGAGCGTTCCACCGCTCCACCTCTAAATCAGTT 1274
Qy 516 hrGln-----ArgCysSerSerLysGlyL 524
Db 1275 GTTCCACTTAAATGACTTCAATCGACAAACCAACGATAGTGTCTTCTTAAACAA 1334
Qy 524 yThrAlaGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysTyrGlyGly 544
Db 1335 GG-----AAAAATCATCGCTGTGTCAAG--TCGTTGTGATATG 1371
Qy 544 LuSerThrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetG 564
Db 1372 AGCAGTCG-----TCCG 1383

```

```

QY 564 luThrGluAsnSerValLeuSerHis---SerAlaLysValSerPro----- 578
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1384 CGTCTGAAGACTCCATTGGCTGCTATCGCTCGGCTCAGTGAAGCTCGCGGACAAAACCTT 1443
QY 579 --AlaGluHisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProIlyshL 598
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1444 CTGGTAATCATTCGCTG-----GAGAGAGGATGGGAAAGATA 1482
QY 598 ysLysLysGlnLysLeuGluValThr-----A 607
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1483 AGACATCAGATCCAGCGGCTACACTCTGAGCGCGGTGTGGCGATGTGCGCCAAATGA 1542
QY 607 rgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysA 627
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1543 CGGAGAAGCTGAAGATACATGATGATCATCTCGT-----CGAGCACAGA 1587
QY 627 snGlnHis-----GluArgGlnLeuMetThrGluThrAspCysSerAspIleA 643
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1588 ACGGTATCTCGGACAACTTCGAGACAGTTCCTCTGCTCGGAATATCCGATACA 1647
QY 643 snArgIleGlnSerLysThrThrAlaAsp-----AspAspCysValIleValAlaA 660
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1648 ACGAGCTCGAGACATATCCAGGAGCATTTGTCCGGAGTAGACATGCAACAGTCGCT 1707
QY 660 laLysAspGlySerAspTyrAlaSerServal---PheAspThrAsnSerGlnGlnLys- 678
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1708 CCAAAACAT---AGCGACTATTCCTCCACTTTGTTCGCCATCCCAAGCTTCCTCCTCAAAGC 1764
QY 679 -----SerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaL 695
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1765 CCGGAGTCCCGAGTCGCTCCATCAGTCAGTTCCTGATCTCGACGAGAA----- 1817
QY 695 euThrThrGlnGluSerProHisProGlnAsnPheGlnSerThrGlnGlnGlnThrH 715
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1818 -----CAGGAGAATGTGTACAAACTCTGTCCAGTCGCGAAGCAAGCAAGTCGCG- 1868
QY 715 isLeuArgMetGluGlnMetValThrIleAlaAlaSerProLeuPheSerHisHis- 734
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1869 -----GCGCTGCCCACTTCACCTTCGACACAACTT 1899
QY 735 -----AspAspGlnTyrIleAlaGluAlaProThrGluHisIleArgLysAspA 752
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1900 CGCTAAGATCCCGGATCACTCATCTATCTCCA-----CACTTATCA---GTGTCAG 1950
QY 752 laLysLysLeuThrTrpGluGlnPheLysAlaThrThrArg-----AsnSerProAlaA 770
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1951 CTGATAGGACACAAATGCTATGCACTCACAGCTAGTCGACGACCTTCCTCAAAAAC 2010
QY 770 laThrCysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisV 790
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2011 CAAGCTATTCAGGCCAATTT-----CATTCATTGAT---CGTAAATGCCACC 2055
QY 790 alMetGlySerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAspArgT 810
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2056 TTCAAGAGTTCAATCCACCGGACGACAGAAATGGCGCTCTCTGAGCCCG----- 2105
QY 810 yrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrM 830
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2106 -----AGCGGTGCCGAACTCGATGTGGAAT 2133
QY 830 etGluAlaSerLysLeuCysAspArgArgAsnAlaGly-----GlnValValLeuTyrP 848
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2134 ATGATTCTTCAGGATCTACTCGGCGCTTCGCGAGTGGAGCTCTACTGCTATCTATG 2193
QY 848 roLysGluSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaA 868
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2194 GAGAG-----ACGTTCCAACTGCACAGACTATCCGATGAAATAATCCCGGAC 2241
QY 868 erPheProAsnTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerG 888
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2242 AT-----TCGCCAAAAGTGAGATGGATCCCACTA----- 2273
QY 888 lnTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsnG 908

```

```

Db 2274 -----TCACTGGCTAGCAGCAGCATATGGATCT---CTCAATG 2310
QY 908 lyLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgP 928
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2311 AGAAGTAC-----GAACATGCTATTTCGGGACATGGCAGCTG 2346
QY 928 roLeu-----ArgProHisProArgValGlyV 937
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2347 ACTTGGAGTGTACAGAACACTGTGCACCTACCTAACCAAGAACAGAGAACTATGGAG 2406
QY 937 alLeuGlySerLeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnS 957
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2407 CAITGTTGATCTTTTGGCAAAGCTTAGAAAACCTCACTCAACACATTCATGATCCCA 2466
QY 957 er-----GlyTyrLysLeuGlyValS 964
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2467 ACTTGAAGCCTGAAGAGGCAATACGATTACGGCGAGGACATTGCTCATTTGAGGGATATA 2526
QY 964 erThrGlyIleThrSerHisGlnMetAsnArgLysGlu----- 976
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2527 GCAATCATCTTGATCCCAACTCAGCTCATGCTAACGAAGGCGCTGTGTGAGCTTCTTCTC 2586
QY 977 -----HisPheGluAlaLeuAsnSerGlyMetPheSerA 988
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2587 AACCATCTCTGGAATCAGTTGCATCCCATCGATCATGATCATGCTGTCTGCGAAAAGCA 2646
QY 988 laLysTrpAsnAlaLeuGlnLeuGlySerValSerSerAlaAspPheLeuSerAlaA 1008
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2647 GCAAGCAGGAGAGATCAGCTTGAGCTCGTTGGCAAGAAC----- 2687
QY 1008 rgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHisProLeuAspA 1028
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2688 -----AAGAAGAGCTGGATCCGCTCCCTCA-----CTCTCCA 2718
QY 1028 rgPheValArgGlnAspIleCysIleThrAsnLysAsn-----ProAlaAspPheThrT 1046
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2719 AGTTCACCAAGAG-----AAGAACAGAGAACTAGCAGCAAGCATATGCCAT 2766
QY 1046 hrIleSerAsnAspAsnGluTyrMetAsp 1055
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 2767 CAATTTCCGGATCTCAAGGAAGCTCTTGAC 2795

```

Search completed: May 17, 2004, 10:07:18  
Job time : 499 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 17, 2004, 08:43:42 : Search time 2543 Seconds  
(without alignments)  
1886.244 Million cell updates/sec

Title: US-09-721-114-2

Perfect score: 5526

Sequence: 1 MBIVADQGGARVGVGTNCML.....NKNPADFTTISNDNEIMDYR 1057

Scoring table:

BLOSUM62 Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2947324 seqs, 2269024515 residues

Total number of hits satisfying chosen parameters: 5894648

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09721114/runat\_14052004\_092255\_4208/app\_query.fasta\_1.1223  
-DB=Published Applications NA -QFWT=fastcap -SUFFIX=rnbp -MINMATCH=0.1  
-LOPCPI=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -LOCAL=LOCAL -OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09721114@cgn2\_1\_693@runat\_14052004\_092255\_4208  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2:  
14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	5526	100.0	3896	9	US-09-828-068-1	Sequence 1, Appli
2	5526	100.0	4310	15	US-10-447-135-1	Sequence 1, Appli
3	5191	93.9	9455	15	US-10-447-135-3	Sequence 3, Appli
4	491.5	8.9	871	13	US-10-425-114-4042	Sequence 4042, Ap
5	223	4.0	2812	15	US-10-028-386-22571	Sequence 22571, A
6	223	4.0	11167	15	US-10-252-157-188	Sequence 188, App
7	222.5	4.0	279	9	US-09-294-093B-2759	Sequence 2759, App
8	220	4.0	3953	15	US-09-864-761-19041	Sequence 19041, A
9	220	4.0	3953	15	US-10-029-386-20602	Sequence 20602, A
10	220	4.0	4383	15	US-10-295-027-427	Sequence 427, App
11	219	4.0	6745	15	US-10-120-988-286	Sequence 286, App
12	214	3.9	4584	10	US-09-945-917A-51	Sequence 51, Appl
13	214	3.9	4584	10	US-09-845-917A-51	Sequence 51, Appl
14	214	3.9	4749	10	US-09-945-917-55	Sequence 55, Appl
15	214	3.9	4749	10	US-09-845-917A-55	Sequence 55, Appl
16	214	3.9	5009	10	US-09-945-917-52	Sequence 52, Appl
17	214	3.9	5009	10	US-09-845-917A-52	Sequence 52, Appl
18	214	3.9	5010	10	US-09-945-917-56	Sequence 56, Appl
19	214	3.9	5010	10	US-09-845-917A-56	Sequence 56, Appl
20	214	3.9	5072	10	US-09-945-917-2	Sequence 2, Appli
21	214	3.9	5072	10	US-09-845-917A-2	Sequence 2, Appli
22	214	3.9	5073	10	US-09-945-917-1	Sequence 1, Appli
23	214	3.9	5073	10	US-09-845-917A-1	Sequence 1, Appli
24	214	3.9	9642	10	US-09-945-917-30	Sequence 30, Appl
25	214	3.9	9642	10	US-09-845-917A-30	Sequence 30, Appl
26	214	3.9	10443	10	US-09-945-917-25	Sequence 25, Appl
27	214	3.9	10443	10	US-09-845-917A-25	Sequence 25, Appl
28	214	3.9	13414	10	US-09-945-917-27	Sequence 27, Appl
29	214	3.9	13414	10	US-09-845-917A-27	Sequence 27, Appl
30	213	3.9	2290	15	US-10-102-524-1741	Sequence 1741, Ap
31	213	3.9	8733	15	US-10-144-198-11	Sequence 11, Appl
32	212.5	3.8	8816	13	US-10-282-122A-7740	Sequence 7740, Ap
33	211	3.8	7474	10	US-09-945-917-26	Sequence 26, Appl
34	211	3.8	7474	10	US-09-845-917A-26	Sequence 26, Appl
35	209	3.8	6852	15	US-10-172-502-3	Sequence 3, Appli
36	207	3.7	5495	13	US-10-282-122A-34849	Sequence 34849, A
37	205	3.7	4217	10	US-09-945-917-49	Sequence 49, Appl
38	205	3.7	4217	10	US-09-845-917A-49	Sequence 49, Appl
39	205	3.7	4382	10	US-09-945-917-53	Sequence 53, Appl
40	205	3.7	4382	10	US-09-845-917A-53	Sequence 53, Appl
41	205	3.7	4642	10	US-09-945-917-50	Sequence 50, Appl
42	205	3.7	4642	10	US-09-845-917A-50	Sequence 50, Appl
43	205	3.7	4643	10	US-09-945-917-54	Sequence 54, Appl
44	205	3.7	4643	10	US-09-845-917A-54	Sequence 54, Appl
45	203	3.7	10567	16	US-10-144-194A-61	Sequence 61, Appl

ALIGNMENTS

RESULT 1

US-09-828-068-1  
Sequence 1, Application US/09828068  
Patent No. US20020157137A1

GENERAL INFORMATION:

APPLICANT: Moon, Yong-Hwan

APPLICANT: Chen, Lingling

APPLICANT: Sung, Zinmay R.

TITLE OF INVENTION: NUCLEIC ACIDS THAT CONTROL REPRODUCTIVE DEVELOPMENT IN

TITLE OF INVENTION: PLANTS

FILE REFERENCE: 018941-001400US

CURRENT APPLICATION NUMBER: US/09/828,068

CURRENT FILING DATE: 2002-04-11

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 3896

TIPS: DNA

ORGANISM: Oryza sativa

FEATURE:

NAME/KEY: CDS

LOCATION: (241)..(3411)

US-09-828-068-1

## Alignment Scores:

Pred. No.: 0 Length: 3896  
 Score: 5526.00 Matches: 1057  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-721-114-2 (1-1057) x US-09-828-068-1 (1-3896)

QY 1 MetGluIleValAlaValAspGlnGluGlyAlaArgValValGlyThrAsnCysMetLeu 20  
 DB 241 ATGGAGATTGTTCCAGTAGATCAGGAGGAGCTCGTGTGTTGGACCAACTGTATGCTT 300  
 QY 21 AlaArgGlyThrGluValAlaProValLeuThrAlaThrProArgGln 40  
 DB 301 GCTGTGTGGAACTGTGCTGTAGCCGACGTGTGGAGCTGACAGCGCGCTCGTCAG 360  
 QY 41 AspAlaAlaAlaGluAlaGlyValAspGluProAlaGlnHisGlnCysGluHisPheSer 60  
 DB 361 GATGACGCGCTGCAAGCTGTGTAGACGAAACCGGACCAACACCAATGCGAGCATTTCTCC 420  
 QY 61 IleArgGlyThrValAlaLeuLeuGlnLysLysAspProLysPheCysSerLeuSerArg 80  
 DB 421 ATAAGAGGGTATGTTGCTCTTCTTCCAGAAAGAGATCCAAATTCGTCTCTATCTCG 480  
 QY 81 IlePheHisAspGlnLysCysAspGluHisLysAlaSerSerProPheSerVal 100  
 DB 481 ATTTTCCATGACCAAGAAAATGTGATGAACACAAAGCTAGTTCAAGCCCAATTTCTGTA 540  
 QY 101 AlaLysPheArgThrPheValProLysCysSerLysCysLeuAspLysLeuLysThrSerAspAsn 120  
 DB 541 GCAAAATTTCCGACGATGGGATTTGCTCGAAGTGCCTGGATTAAGTTGAAACTTCAGATAAT 600  
 QY 121 GlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIle 140  
 DB 601 GGAACAGCACCAAGAACTCTTCCGCAAGCAGATGCGACAAAGTGTGTTGCTCCATC 660  
 QY 141 ThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysValSerProSer 160  
 DB 661 ACATTTGTTCCGAGCATTGTTGCTGCTAGTGTGTTGTTGCTCCAAAGAGTGTCTCTAGC 720  
 QY 161 ThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGlnGlu 180  
 DB 721 ACACATCATCTCAAGGGAAGATGCTGATGATCACTCTTCCAAGAGTGTGCAAGAA 780  
 QY 181 GlyAsnAspSerLysCysAsnAlaProSerGlyLysAsnGlyAlaAlaGluAlaSerThr 200  
 DB 781 GGCATATGACTCCAAATGCAATGCGCTTCTGCAAGAAATGGAGCTGCTGAGGCCAATACT 840  
 QY 201 AspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAlaAlaAsnVal 220  
 DB 841 GATTCACCAATGAAGAATTTCAGAGGCGCGCCCAAAATATGATGTGCGACCAATGTC 900  
 QY 221 SerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIleThrTrp 240  
 DB 901 TCTGAGGACACACTTCTGTTGATGTGGGCTTTTACCTCCATCCCTCCAAACTTCTGAAAGTGGTC 960  
 QY 241 HisIleGluValAsnGlyAlaAspGlnProProSerThrProLysLeuSerGluValVal 260  
 DB 961 CACATAGAGTAATGTTGGTCAGATCAACCTCATCCCTCCAACTTCTGAAAGTGGTC 1020  
 QY 261 LeuLysArgAsnGluAspGluAsnGlyLysThrGluGluThrLeuValAlaGluGlnCys 280  
 DB 1021 CTCAAAAGAAATGAAGTGAATGGAAAACCTGAAGAGACTTTGTTGCTGAGCAGTGC 1080  
 QY 281 AsnLeuThrLysAspProAsnProMetSerGlyLysGluArgAspGlnValAlaGluGln 300  
 DB 1081 AATTTGACCAAGATCTCAACCAATGTTCTGGAAGAGAACTGATCAGGTTGCTGAGCAG 1140  
 QY 301 CysAsnLeuThrLysAspProLysProValSerGlyLysGlnLysCysGluGlnIleCysAsn 320

DB 1141 TGCAATTTGACCAAAAGATCCGAAACCAAGTGTCTGGCGAGAAATCTGACGAGATCTGCAAT 1200  
 QY 321 GluProCysGluGluValValLeuLysArgSerSerLysSerLysArgLysThrAspLys 340  
 DB 1201 GAGCATGTGGAAGAAGTTGTTCTCAAGAAGAGCTCCAAATCTAAGAGGAGACGCGATAG 1260  
 QY 341 LysLeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAsp 360  
 DB 1261 AAGTTGATGAAGAAGCAGCAGCAGCAAGAAACGCACTGCCAGGCTGATGTTTCAGAT 1320  
 QY 361 AlalysLeuCysArgArgLysProLysValArgLeuLeuSerGluLulleIleAsnAla 380  
 DB 1321 GCAAGCTTTGTCGAGAAAGCCAAAAGAGTGGCGCTTCTATCAGAAATATATAATGCT 1380  
 QY 381 AsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspProCys 400  
 DB 1381 AACCAAGTTGAGGATTTCTAGAAGTGACGAGCTTCATCGTGAATCGGCTGATCCCTGT 1440  
 QY 401 GluAspArgSerThrIleProValProMetGluValSerMetAspIleProValSer 420  
 DB 1441 GAGATGATGAAGTACCATCCCGTCCCGATGGAAGTAAAGCATGGATATTTCTGTTAGC 1500  
 QY 421 AsnHisThrValGlyGluAspGlyLeuLysSerSerLysAsnLysThrLysArgLysTrp 440  
 DB 1501 AACCATACAGTGGAGAGATGGGTAAATCAAGTAAGAACAGACAAACGCAATAC 1560  
 QY 441 SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysArg 460  
 DB 1561 TCTGATGTTGTAGATGATGATCATCTTATGAATCGGCTGATGGTGAATGGAAAAGAAAGA 1620  
 QY 461 ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
 DB 1621 ACTGGAAGTGTGCATCACAGTTGCTTCATCCAGCTGGAAATTTGACGACAAACAAAGTG 1680  
 QY 481 ThrProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAspThr 500  
 DB 1681 ACACCACTGCGAGTACTCAGCATGATGATGAGTAATGATGATGATACTGATAATGGCTTGACACA 1740  
 QY 501 AsnMetHisLysThrAspValCysGlnHisValSerGluLeuSerThrGlnArgCysSer 520  
 DB 1741 AATATGATGAAGACAGATGTTGTGACGATGTATCAGAAATCTCCACAGAGAGTGCTCA 1800  
 QY 521 SerLysGlyLysThrAlaGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLys 540  
 DB 1801 TCAAAGGGGAAACAGCGGGTTTGAGTAAGGGGAAACACATTCAGCTGCTAGTACCAAA 1860  
 QY 541 TyrGlyGlyGluSerThrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGln 560  
 DB 1861 TATGGTGTGAAGACACCGAAATGCGTCAGAACATACATGTTACTCAGCGCAGAGATCAA 1920  
 QY 561 CysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLysValSerProAlaGlu 580  
 DB 1921 TGCCAGATGGAAACCGAAACTCTGTTCTGAGTCACCTCGGCAAGGTTTCTCCAGCTGAG 1980  
 QY 581 HisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLys 600  
 DB 1981 CATGATATCCAAATATATGCTGACCTTTCATGAGCAGAGTCTACCCAAAGAAAGAAAG 2040  
 QY 601 GlnLysLeuGluValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIle 620  
 DB 2041 CAANAATCTGAGTACTCTGTGAANAACACACCATGATGATGATGCCATCCCATGATATT 2100  
 QY 621 ValGluLeuAlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSer 640  
 DB 2101 GTTCAACTGTAGTAAACCAACAGCATGAGCAGCGCTTATGACTGAGACTGATTTGTTCT 2160  
 QY 641 AspIleAsnArgIleGlnSerLysThrAlaAspAspCysValIleValAlaAla 660  
 DB 2161 GACATCAACGATATTCAATCCAGACAACTGCTGATGATGATGATGATGATGATGATGAT 2220  
 QY 661 LysAspGlySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGlnLysSerLeu 680  
 DB 2221 AAGGATGTTTCAGATTATGATCAATGAAGTGTGTTGACACTAATTTCCCAACAGAGTCTGTT 2280

QY 681 AlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGluSer 700  
DB GCATCCCAAGATACAGAGAGAGTACAGGGTCAATTTGGCATTTGACCAACAGAGTCT 2340  
QY 701 ProHisProGlnAsnPheGlnSerThrGlnGlnGlnThrHisLeuArgMetGluGlu 720  
DB CCATCCCTCAGAACTTTCAGTCTACTCAGGAACAGCAGACATTTGGGATGGAGAA 2400  
QY 721 MetValThrIleAlaAlaSerSerProLeuPheSerHisHisAspGlnTyrIleAla 740  
DB ATGGTCACATATTGCTGCAAGCTCACCACATTTTTCATCATCATCATCATCATATATGCT 2460  
QY 741 GluAlaProThrGluHisTrpGlyArgLysAspAlaLysLysLeuThrTrpGluGlnPhe 760  
DB GAAGCACCACCTGAACATTTGGGCGGTAGGAGCGCAAGAGCTAAGTGGGAGCAATTT 2520  
QY 761 LysAlaThrThrArgAsnSerProAlaAlaThrCysGlyAlaGlnPheArgProGlyIle 780  
DB AAGGCCACTACAGAAATTTCTCCAGCAGCAACATGTGGTCTCAATTTAGACCTGGTATC 2580  
QY 781 GlnAlaValAspLeuThrSerThrHisValMetGlySerSerSerAsnTyrAlaSerArg 800  
DB CAAGCAGTTGACTTGACTTCTACTCATGTCTATGGATCTTCCAGCAATTTATGCATCTCC 2640  
QY 801 GlnProValIleAlaProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAla 820  
DB CAACCACTAATTCGGCCACTGGACCGCTATGCTGAAGAGCGGTAAACCAAGTCCATGCA 2700  
QY 821 ArgAsnPheProSerThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsn 840  
DB AGAAATTTTCCAGCACAATAGCAACCATGGAAGCGAGTAAGTATGTATCGAGAAAT 2760  
QY 841 AlaGlyGlnValValLeuTyrProLysGluSerMetProAlaThrHisLeuLeuArgMet 860  
DB GCTGGACAAGTACTGTGTATCTCTAAGAATCATGCTCGCAGCATCTCTCGAAGATG 2820  
QY 861 MetAspProSerThrLeuAlaSerPheProAsnTyrGlyThrSerSerArgGlnMet 880  
DB ATGGATCCATCAACATTAGCAGCTTCCCACTATGGAACCTTCTAGCAGGACCAAGATG 2880  
QY 881 GluSerGlnLeuHisAsnSerGlnTyrAlaHisAsnGlnTyrLysGlySerThrSerThr 900  
DB GAGCTCTCACTTCATATTTCTCAGTATGCACATATCAGTACAAAGATCAACACGACCA 2940  
QY 901 SerTyrGlySerAsnLeuAsnGlyLysIleProLeuThrPheGluAspLeuSerArgHis 920  
DB TCATATGGCAGTAACCTGAATGGAAGATTTCCATTGACATTTGAAAGACTTATCACCGCAT 3000  
QY 921 GlnLeuHisAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlySer 940  
DB CAGCTGCATGATCTGCACAGACCTTTACGCCACATCTTAGAGTTGGTGTGCTTCC 3060  
QY 941 LeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLys 960  
DB TTGCTCGAGAAGGAAATTCGAAACTGGTCGAGAACTGTGGCACAATCTGTTATATAAG 3120  
QY 961 LeuGlyValSerThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAla 980  
DB TTAGGAGTGTCAACAGATATACATCGCATCAGATGACAGAAAGAAATTTTGAAGCC 3180  
QY 981 LeuAsnSerGlyMetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSer 1000  
DB CTGAATTTCTGGAATGTTTTCAGCAAAATGAAATGCAATTCAGTTGGGTTCTGTAGCTCC 3240  
QY 1001 SerAlaAspPheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgLysGly 1020  
DB AGTGCAGATTTTATTCAGCGAGGAACAGCATAGCTCAATCTTGGACCAAGAGCGAGGT 3300  
QY 1021 LysMetValHisProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAsn 1040  
DB AAAATGGTTCATCCCTTGGATCGTTTGTGAGACAGATATCTGTATATACACAGAAC 3360

QY 1041 ProAlaAspPheThrThrIleSerAsnAspAsnGluTyrMetAspTyrArg 1057  
DB CCAGCTGATTTTACTACATCATCATCATCATCATCATCATCATCATCATCATCATCAT 3411  
RESULT 2  
US-10-447-135-1  
; Sequence 1, Application US/10447135  
; Publication No. US20030199684A1  
; GENERAL INFORMATION:  
; APPLICANT: Yamazaki, Hirohiko  
; APPLICANT: Miyao, Akio  
; TITLE OF INVENTION: A novel gene involved in brassinosteroid responses  
; FILE REFERENCE: MAPF-1 DIV  
; CURRENT APPLICATION NUMBER: US/10/447,135  
; CURRENT FILING DATE: 2003-05-27  
; PRIOR APPLICATION NUMBER: US 09/721,114  
; PRIOR FILING DATE: 2000-11-22  
; PRIOR APPLICATION NUMBER: JAPAN 2000-149106  
; PRIOR FILING DATE: 2000-05-19  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4310  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (655)..(3828)  
US-10-447-135-1  
Alignment Scores:  
Pred. No.: 0 Length: 4310  
Score: 5526.00 Matches: 1057  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 15 Gaps: 0  
US-09-721-114-2 (1-1057) x US-10-447-135-1 (1-4310)  
QY 1 MetGluIleValAlaValAspGlnGluGlyAlaArgValAlaGlyThrAsnCysMetLeu 20  
DB 655 ATGGAGATGTTGTCAGTAGATCAGGAGGAGCTGTGTTGTTGGCAACATGTATGTT 714  
QY 21 AlaArgGlyGlyThrGlyAlaValAlaProValLeuGluLeuThrAlaThrProArgGln 40  
DB 715 GCTCTGTGTGAACCTGGTGTGTAGCCGACGTGTTGGAGCTGACAGCGCTCTGTGAG 774  
QY 41 AspAlaAlaAlaGluAlaGlyValAspGluProAlaGlnHisGlnCysGluHisPheSer 60  
DB 775 GATGACGCGCTGAGCTGTGTAGACGACCGGACCAACCAATGCGAGCATTTCTCC 834  
QY 61 IleArgGlyTyrValAlaLeuLeuGlnLysLysAspProLysPheCysSerLeuSerArg 80  
DB 835 ATAGAGGATGATGTTGCTCTTCTTCAGAAAGAGGATCCAAAATTCGTCTCTATCTCG 894  
QY 81 IlePheHisAspGlnLysLysCysAspGluHisLysAlaSerSerSerProPheSerVal 100  
DB 895 ATTTTCCATGACCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 954  
QY 101 AlaLysPheArgArgTrpAspCysSerLysCysLeuAspLysLeuLysSerSerAspAsn 120  
DB 955 GCAAAGTTTCCAGCATGCGATTCCTCGAAGTCTTGGATTAAGTTGAAACCTCAGATAAT 1014  
QY 121 GlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIle 140  
DB 1015 GCACAGCACCAGAAACTCTTCCCCCAAGCAGAAATGCAAGATGATGATGATGATGATG 1074  
QY 141 ThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysValSerProSer 160  
DB 1075 ACAATTTGTTGGAGCACATTTTGTGCTGTAGTGTGTTGTTGTTGTTGTTGTTGTTG 1134

161 ThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGlnGlu 180  
1135 ACACAATCATCTCAAGGGAAGAAATGCTGATAGATCAACTCTTCCAAAGAGTGTGCAGAA 1194  
181 GlyAsnAspSerLysCysAsnAlaProSerGlyLysAsnGlyValAlaAlaAsnThr 200  
1195 GCGATGACTCCAAATGCAATCGGCTTCTGCGAAGATGGAGCTGCTGAGGCCAATACT 1254  
201 AspSerProMetLysAspLeuGlnGlyProAlaGlnAsnThrAspValAlaAlaAsnVal 220  
1255 GATTCCACCAATGAAGAATTGCAAGGGCCAGCCCAAAATATATGATGTGCGACCAATGTC 1314  
221 SerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIleThrTrp 240  
1315 TGTGAGGCAACACTTCTGTGATGTGGGCTTTACCTGAAGTTCCCGACGATTACTAGG 1374  
241 HisIleGluValAsnGlyAlaAspGlnProProSerThrProLysLeuSerGluValVal 260  
1375 CACATAGAAGTAAATGGTGCAGATCAACCTCCATCCACTCCAAAATTTCTGAAAGTGTCT 1434  
261 LeuLysArgAsnGluAspGluAsnGlySerThrGluGluThrLeuValAlaGluGlnCys 280  
1435 CTCAAAGAAATGAAGATGAAATGGAAAACTGAAGACACTCTGTGCTGAGCAGTGC 1494  
281 AsnLeuThrLysAspProAsnProMetSerGlyLysGluArgAspGlnValAlaGluGln 300  
1495 AATTTGACCAAGATCTCTAACCAATGTCTGGAAGGAAGTGCAGGTTGCTGAGCAG 1554  
301 CysAsnLeuThrLysAspProLysProValSerGlyGlnLysCysGluGlnIleCysAsn 320  
1555 TGCAAATTTGACCAAGATCTCTAACCAATGTCTGGAAGGAAGTGCAGGTTGCTGAGCAG 1614  
321 GluProCysGluGluValValLeuLysArgSerSerLysSerLysArgLysThrAspLys 340  
1615 GACCCATGTGAAGAGTGTCTCAAGAGAGCTCCAAATCTAAGAGGAAGCAGGATAG 1674  
341 LysLeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAsp 360  
1675 AAGTTGATGAAGAAGCAGCAGCAGCAGCAAGAAAGCAGCTGCCAGGCTGATGTTTCAGAT 1734  
361 AlaLysLeuCysArgArgLysProLysLysValArgLeuLeuSerGluIleIleAsnAla 380  
1735 GCAAGCTTTGTCCGGAAGGCGAAAGAGTGGCTGCTCTATCAGAAATTTAATAGT 1794  
381 AsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspProCys 400  
1795 AACCAAGTTGAGGATCTAGAGTGCAGAGTTCATCGTCAAAATGCCGCTGATCCCTGT 1854  
401 GluAspAspArgSerThrIleProValProMetGluValSerMetAspIleProValSer 420  
1855 GAGGATGATAGAAGTACCATCCCGGCTCCCGATGGAAGTAAAGTATCTCTGTAGC 1914  
421 AsnHisThrValGlyLysAspGlyLeuLysSerSerLysAsnLysThrLysArgLysThr 440  
1915 AACCATACAGTGGGAGAGATGGTTAAATCAAGTAAGAAACAAGACAAACGCAAAATAC 1974  
441 SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysLysArg 460  
1975 TCTGATGTTGATGATGATCATCATCTATGAACTGGCTGAATGGAAAGAAAGAAAGA 2034  
461 ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
2035 ACTGGAAGTGTGCATCACAGATTGCTCATCCAGCTGGGAATTTGACCAACAAAAAGTG 2094  
481 ThrProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAspThr 500  
2095 ACACCCATCGGAGTACTACGACATGATGATGAGATGATGATGATAATGTTCTGTGACAC 2154  
501 AsnMetHisLysThrAspValCysGlnHisValSerGluIleSerThrGlnArgCysSer 520  
2155 AATATGATATACAGAGTGTCTGTGACATGTATCAGAAATCTCCACAGAGGTGCTCA 2214  
521 SerLysGlyLysThrAlaGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLys 540

2215 TCAAAGGGGAANAACAGCGGTTTGAAGTAAGGGGAANAACACATTCAGCTGCTAGTACCAA 2274  
541 TyrGlyGlyGluSerThrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAsnGln 560  
2275 TATGGTGTGAAAGACACAGAAATGGTCAGACATACATGCTACTCAGCGCAGAGATCAA 2334  
561 CysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLysValSerProAlaGlu 580  
2335 TGCAGATGGAAACCGAAACTCTGTTCTGAGTCACTCGGCAAAAGGTTTCTCCAGCTCAG 2394  
581 HisAspIleGlnIleMetSerAspLeuHisGlnSerLeuProLysLysLysLysLys 600  
2395 CATCATATCCAAATATATGTTCTGACCTTTCATGAGCAGAGTCTTACCCAAAGAAAGAAAG 2454  
601 GlnLysLeuGluValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIle 620  
2455 CAAAACCTTGAAGTGATCTGTGAAANAACGACCATGATAGATGACATCCCATGGAATTT 2514  
621 ValGluLeuLeuAlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSer 640  
2515 GTTGAACCTGCTAGCTAAANAACGACCATGAGGAGGCTTATGACTGAGACTGATTTGTCT 2574  
641 AspIleAsnArgIleGlnSerLysThrAlaAspAspAspCysValIleValAlaAla 660  
2575 GACATCAACCGTATTCATCCACACACTCTCTGATGATGATGATGATGATGATGATGATG 2634  
661 LysAspGlySerAspTyrAlaSerSerValPheAspThrAsnSerGlnGlnLysSerLeu 680  
2635 AAGCATGGTTCCAGATATGTCATCAAGTGTGTTGCACACTAATTCCTCCACAGAGTCTTG 2694  
681 AlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThrThrGlnGluSer 700  
2695 GCATCCCCAAGATACAGAGAGGTTACAGGCTCATTTGGCATTTGACCAACACAGAGTCT 2754  
701 ProHisProGlnAsnPheGlnSerThrGlnGlnGlnThrHisLeuArgMetGluGlu 720  
2755 CCACATCTCTGAACTTTCAGTCTACTCAGGAACAGCAGACACACATTTCCGCGTAGAAGAA 2814  
721 MetValThrIleAlaAlaSerSerProLeuPheSerHisHisAspAspGlnTyrIleAla 740  
2815 ATGGTCACTATTGCTGCAAGCTCACCCTATTTTCCATCATCATGATGATGATGATGATG 2874  
741 GluAlaProThrGluHisTrpGlyArgLysAspAlaLysLysLeuThrTrpGluGlnPhe 760  
2875 GAAGCCCACTGAACAATGGGCGCTAAGAGCGCAAGAGCTAACCTGGGAGCAATTT 2934  
761 LysAlaThrThrArgAsnSerProAlaAlaThrCysGlyAlaGlnPheArgProGlyIle 780  
2935 AAGGCCACTACAAAGAAATTTCCAGCAGCAACATGTGGTGTCTCAATTTTAGACCTGATC 2994  
781 GlnAlaValAspLeuThrSerThrHisValMetGlySerSerSerAsnTyrAlaSerArg 800  
2995 CAAGCAGTGTGACTTGACTTCTACTCATGTCTGGAATCTTCCAGCAATTTGCACTCTGC 3054  
801 GlnProValIleAlaProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAla 820  
3055 CAACCCAGTAAATTCGCCACTGCAACCGCTATGCTGAAGAGCGGTTAACCGATCCATGCA 3114  
821 ArgAsnPheProSerThrIleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsn 840  
3115 AGAAATTTTCCAAAGCACAAATAGCAACCTGGAAGCGAGTAGTATGATGATGATGATGAT 3174  
841 AlaGlyGlnValValLeuTyrProLysGluSerMetProAlaThrHisLeuLeuArgMet 860  
3175 GTTGGACAAAGTATGTTCTGTATCTTAAGAAATCCATGCTCGCAGCAGCATCTTCTGAGAATG 3234  
861 MetAspProSerThrLeuAlaSerPheProAsnTyrGlyThrSerSerArgAsnGlnMet 880  
3235 ATGGATCCATCAACATTAGCAAGCTTCCCAACTATGGAACCTTCTAGCAGGAACCAAGTG 3294  
881 GluSerGlnLeuHisAsnSerGlnTyrAlaHisAsnGlnTyrLysGlySerThrSerThr 900

Db 3295 GAGTCTCAACTTCATAATCTCAGTATGCACATAATCAGTACAAAGGATCAACCCAGCACA 3354  
Qy SerTyrGlySerAsnLeuAsnGlyLysIleProLeuThrPheGluAspLeuSerArgHis 920  
Db 3355 TCATATGCGAGTAACTGAATGGAAGATTCATTCATTCAGACTTCATCAGGCAT 3414  
Qy 921 GlnLeuHisAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlySer 940  
Db 3415 CAGCTGCATGATCTGCACAGACCTTTACGCCACATCTCAGTGGTGGCTCC 3474  
Qy 941 LeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyLys 960  
Db 3475 TTGCTGCAGAGGAAATTCGAATCTGTCGAGNACTGTGCGCACCAATCTGGTTATPAG 3534  
Qy 961 LeuGlyValSerThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAla 980  
Db 3535 TTAGGAGTGTCAACAGCAATAACATCGCATCAGATGAACAGAAAGAACATTTTGAAGCC 3594  
Qy 981 LeuAsnSerGlyMetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSer 1000  
Db 3595 CTGAATCTCGAATGTTTTCAGCAAAATGAATGCATTCGATGGGTTCTGTAGTCC 3654  
Qy 1001 SerAlaAspPheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGly 1020  
Db 3655 AGTGCAGATTTTATCAGCGAGGAACAGCATAGCTCAATCTGGACCAAGGAGGT 3714  
Qy 1021 LysMetValHisProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAsn 1040  
Db 3715 AAAATGTTTCATCCCTTGGATCGGTTTGTGAGACAGGATATCTGTATTAACCTAACAGAAC 3774  
Qy 1041 ProAlaAspPheThrThrIleSerAsnAspAsnGlnLysMetAspTyrArg 1057  
Db 3775 CCAGCTGATTTTACTACAACTCAGTAACGATACGATATATGATTAACCGC 3825

## RESULT 3

US-10-447-135-3  
; Sequence 3, Application US/10447135  
; Publication No. US200301996841  
; GENERAL INFORMATION:  
; APPLICANT: Hirochika, Hirohiko  
; APPLICANT: Yamazaki, Muneko  
; TITLE OF INVENTION: A novel gene involved in brassinosteroid responses  
; FILE REFERENCE: MAFF-1 DIV  
; CURRENT APPLICATION NUMBER: US/10/447,135  
; PRIOR FILING DATE: 2003-05-27  
; PRIOR APPLICATION NUMBER: US 09/721,114  
; PRIOR FILING DATE: 2000-11-22  
; PRIOR APPLICATION NUMBER: JAPAN 2000-149106  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 9455  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
US-10-447-135-3

Alignment Scores:  
Pred. No.: 0 Length: 9455  
Score: 5191.00 Matches: 1055  
Percent Similarity: 68.06% Conservative: 0  
Best Local Similarity: 68.06% Mismatches: 2  
Query Match: 93.94% Indels: 495  
DB: 15 Gaps: 3

US-09-721-114-2 (1-1057) x US-10-447-135-3 (1-9455)

Qy 1 MetGluIleValAlaValAspGlnGluGlyAlaArgValValGlyThrAsnCysMetLeu 20  
Db 4223 ATGAGATTTGTTCAGTAGATCAGGAGGAGCTCGTGTTCGCGAGCAACTGTATGCTT 4282  
Qy 21 AlaArgGlyGlyThrGlyAlaValAlaProValLeuGluLeuThrAlaThrProArgGln 40

Db 4283 GCTCGTGTGGAACTGGTGTCTAGCGCCAGTGTGGAGCTGACAGCGCGCTCGTCAG 4342  
Qy 41 AsnAlaAlaGluAlaGlyValAspGluProAlaGlnHisGlnCysGluHisPheSer 60  
Db 4343 GATGCGCGCGCTGAAGCTGGTGTAGACGAACCGGACAAACCAACATCGGAGCATTTCTCC 4402  
Qy 61 Ile----- 61  
Db 4403 AT- AAGGTAATCATTTTCTGTATTTCCAATTCAGTATCGCGTTGTGGATGAATAATGAA 4461  
Qy 61----- 61  
Db 4462 TCGGCATGTCATGCCCATATTCGCACTGTTTGAAGAGAGTATGATTGATCGTGGTTTTTG 4521  
Qy 61----- 61  
Db 4522 CACAGTTTCTGTGGGACTTATATGGTCATCTGTCTTTGTGACGATCGTATACACTGGGTC 4581  
Qy 61----- 61  
Db 4582 GACATGCTTATGCACTTTGGTTCGATTTAGGAAGTCAATACATCCACTACTAGCTCTATAT 4641  
Qy 61----- 61  
Db 4642 CTAGCCATGTGAACCTCAATTTATGCCATAGACAGCTAGCAGGCTAGCAGCAGAAATATA 4701  
Qy 61----- 61  
Db 4702 TATAATATTTGCATATATATGTTGGTGTTCATGCTATCTTTATATCTTACGTACATCCATTA 4761  
Qy 61----- 61  
Db 4762 ATATCTTCAATGTATGAATCTGAGCAGACATGATGTGAGTGTCTACACATATGCGATGCTGT 4821  
Qy 61----- 61  
Db 4822 ATGTGTGTTTCAATAGGTGTTTGCATCATATTTGTTGTGGGGTGGCGATGCAATTTAT 4881  
Qy 61----- 61  
Db 4882 CAGGCCATGCTGTAGGCTGTAGATATTTGTGTGTGTATATTTCTGTGTGACAA 4941  
Qy 61----- 61  
Db 4942 GCTGATTACTAATGAATTAACCTTTTGGGCTACACTCATATATTTGGGCCCTTACATTTT 5001  
Qy 61----- 61  
Db 5002 TCTAATCATTTTTCCTTTGTGCTGAGGTTTCAGCATAAACCTTTTATCATTAAGCATGTT 5061  
Qy 61----- 61  
Db 5062 TACATCCTTAGGAGATTTCTTAGAAGTGTGTTTCTTCATATTTGCAATATGTTTGTATTG 5121  
Qy 61----- 61  
Db 5122 ATAGTCCATTAATTTTAAAGCCTTTTCAATTTGTTTAGAGATTTCTAGAGATGATATAT 5181  
Qy 61----- 61  
Db 5182 CAACCATAGACTGTGCACGTTTTGGTTTAACTTTCTAGAACTAATTAGATTAATTTT 5241  
Qy 61----- 61  
Db 5242 TTGTAGTTTATCCTGTGTATGTTTGTATTTATCTTTTGAATTTCAAACCTGCAATACTAGA 5301  
Qy 61----- 61  
Db 5302 TTATCTTGAAGTCTCTTTTCTTGTGAGCTGTCAAGCTATGTAGAAATGCCTACCTCCC 5361  
Qy 61----- 61



Db 5362 AGCATCTTTAGATTATGTAGGGCCCTTTCTGAGTTTATCAGTTGATATTGACTGGAAGC 5421  
QY 61 ----- 61  
Db 5422 ACGCAATGCTATATATATGTGCCATGCTTTTAAATGATAATCTTTATTTCTTGT 5481  
QY 62 ---- ArgGlyTyrValAlaLeuLeuGlnLysLysAspProLysPheCysSerLeuSerArg 80  
Db 5482 ACCAGAGGGTATGTTGCTCTCTTCAGAGAAGAGGATCCAAAATTTCTCTCTATCTCGG 5541  
QY 81 IlePheHisAspGlnLysLysCysAspGluHisLysAlaSerSerSerProPheSerVal 100  
Db 5542 ATTTTCCATGACCAAGAAAATGTGATGAACAACAAGCTAGTTCAAGCCCATTTCTGTGA 5601  
QY 101 AlaLysPheArgArgTirAspCysSerLysCysLeuAspLysLeuLysThrSerAspAsn 120  
Db 5602 GCAAAATTTTCGACGATGGGATGCTCGAAGTGCTTGGATAAGTTGAAAACCTTCAGATAAT 5661  
QY 121 GlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIle 140  
Db 5662 GGACACGACCCAGAACTCTTCCCGCAAGACGAAATGGCACAAAGTGATGGTTGCTCCATC 5721  
QY 141 ThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysValSerProSer 160  
Db 5722 ACATTTGTTCCGAGCACTTTTGTGCTGCTAGTTGTTGGTTCCCAAAAAGCTGCTCTAGC 5781  
QY 161 ThrGlnSerSerGlnGlyLysAlaAspArgSerThrLeuProLysSerValGlnGlu 180  
Db 5782 ACACAATCATCTCAAGGGAAGATGCTGATAGATCAACTCTTCCAAAGAGTGTGCAAGAA 5841  
QY 181 GlyAsnAspSerLysCysAsnAlaProSerGlyLysAsnGlyAlaAlaGluAlaAsnThr 200  
Db 5842 GGCATGACTCCAAATGCCAATGCCCTTCTGGCAAGATGGAGCTGCTGAGGCCAATACT 5901  
QY 201 AspSerProMetLys----- 205  
Db 5902 GATTCCCAATGAA - AGGTATGTTAGATGATAGAGCCCTTCAAATTCCTAAGTAGGATTT 5960  
QY 206 ----- AspLeuGlnGlyProAla 211  
Db 5961 ATTTAGGTATAGATAAATAATGTTGTGTGATTTTCTCAGATTTGCAAGGGCCAGCC 6020  
QY 212 GlnAsnTyrAspValAlaAlaAsnValSerGluAspAsnThrSerValAspValGlyAla 231  
Db 6021 CAAAATTATGATGGCAGCAAAATGCTCTGAGGACCAACACTCTGTTGATGTTGGGGCT 6080  
QY 232 LeuProGluValProGlnIleThrTirPheIleGluValAsnGlyAlaAspGlnProPro 251  
Db 6081 TTACCTGAAGTTCCCGAGATTACATGCGACATAGAAATGAATGGTGCAGATCAACTCCA 6140  
QY 252 SerThrProLysLeuSerGluValValLeuLysArgAsnGluAspGluAsnGlyLysThr 271  
Db 6141 TCCACTCCAAAACCTTCTGAAGTGGTCTCTCAAAGAAATGAAGATGAAAATGGAAAAACT 6200  
QY 272 GluGluThrLeuValAlaGlnCysAsnIleuThrLysAspProAsnProMetSerGly 291  
Db 6201 GAAGAGACTCTTGTGCTGAGCAGTGCAMTTTGACCAAGATCTCTAACCCAAATGCTGGA 6260  
QY 292 LysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLysProValSer 311  
Db 6261 AAGGAACGTGATCAGGTTGCTGAGCAGTGCATTTTGACCAAGATCCGAAACCAAGTGTCT 6320  
QY 312 GlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluValValLeuLysArgSer 331  
Db 6321 GGCCAGAAATGTGAGCAGATCTGCAATGAGCCATGATGAAGAGTGTGTTCTCAAAAGAAGC 6380  
QY 332 SerLysSerLysArgLysThrAspLysLysLeuMetLysLysGlnGlnHisSerLysLys 351  
Db 6381 TCCAAATCTAAGGAGGAGACGGATAGAAGTTGATGAAGACAGCAGCAGCACAGCAAGAA 6440  
QY 352 ArgThrAlaGlnAlaAspValSerAspAlaLysLeuCysArgArgLysProLysLysVal 371  
Db 6441 CGCACTGCCCGGCTGATGTTTTCAGATGCAAAAGCTTTTGTGGGAGAAAGCCAAAAGAGTG 6500

QY 372 ArgLeuLeuSerGluIleIleAsnAlaAsnGlnValGluAspSerArgSerAspGluVal 391  
Db 6501 CGGCTTCTATCAGAAATATATAATGCTAAACAGGTTTCCAGGATTTAGAGGTACAGAAAT 6560  
QY 392 HisArgGluAsnAlaAlaAspProCysGluAspAspArgSerThrIleProValProMet 411  
Db 6561 CATCGTGAATAATGCGCTGATCCCTGTGAGGATGATAGAAGTACCATCCCGGTCCCGATG 6620  
QY 412 GluValSerMetAspIleProValSerAsnHisThrValGlyGluAspGlyLeuLysSer 431  
Db 6621 GAAGTAAGCATGATATTCCTGTAGCAACCATACAGTGGGAGAGATGGGTTAAAAATCA 6680  
QY 432 SerLysAsnLysThrLysArgLysTyrSerAspValValAspAspGlySerSerLeuMet 451  
Db 6681 AGTAGAACCAAGACAAACGCAATATCTCTGATGTTGTAGTATGATCATCATCTTATG 6740  
QY 452 AsnTirPLeuAsnGlyLysLysLysArgThrGlySerValHisHisThrValAlaHisPro 471  
Db 6741 AACTGGCTGAATGGAAAAAGAAAGAACTGGAAGTGTGCATCACAGTTGCTCATCCA 6800  
QY 472 AlaGlyAsnLeuSerAsnLysLysValThrProThrAlaSerThrGlnHisAspAspGlu 491  
Db 6801 GCTGGAAATTTTGGCAACAAAAAGTAGTACCCACTGCGAGTACTCAGCATGATGATGAG 6860  
QY 492 AsnAspThrGluAsnGlyLeuAspThrAsnMetHisLysThrAspValCysGlnHisVal 511  
Db 6861 AATGATCTGAAATGGTCTTGCACAAATATGCTAAGACAGATGCTCTGTCAAGCATGTA 6920  
QY 512 SerGluIleSerThrGlnArgCysSerSerLysGlyLysThrAlaGlyLeuSerLysGly 531  
Db 6921 TCAGAAATCTCCACACAGAGGTGCTCATCAAGGGGAAACAGCGGGTTTTCAGTAAGGGG 6980  
QY 532 LysThrHisSerAlaAlaSerThrLysTyrGlyGlyGluSerThrArgAsnGlyGlnAsn 551  
Db 6981 AAAACACATTCAGCTGTGTAGTACCAATATGCTGTGTGAAGCACCAAGAAATGGTCAGAAC 7040  
QY 552 IleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrGluAsnSerValLeuSer 571  
Db 7041 ATACATGTACTCAGCGCAGAGATCAATGCCAGATGGAAACCGAAACTCTGTCTGAGT 7100  
QY 572 HisSerAlaLys----- 575  
Db 7101 CACTCGGCAAGAGGTACGAAATTTGTGATCATGAGGAATTTTGTCTTTTAAATGACTG 7160  
QY 575 ----- 575  
Db 7161 AATCAACATTTATCTGTATGAAGCAATAATATGTTGTCATAACAATGTTAAGAAATATGC 7220  
QY 575 ----- 575  
Db 7221 ATACAAATGTTTATTTATGAGCTTTCCACTGTCTCTTCTTACTTATGTTTGTACTCTTT 7280  
QY 575 ----- 575  
Db 7281 TTGTGTGCGGTGCATGTCATGTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 7340  
QY 575 ----- 575  
Db 7341 ACGTGCGTGGCGCAATATCTTTTATAGACTCAATATATAGTATGTTAATGGACTGACA 7400  
QY 576 ----- ValSerProAlaGluHisAspIleGlnIleMetSerAs 588  
Db 7401 TTTTCTCTATTTCTCATCTCAGGTTTCTCCAGCTGAGCATGATATCCAAATATATGCTGA 7460  
QY 588 pLeuHisGluGlnSerLeuProLysLysLysGlnLysLeuGluValThrArgGlu 608  
Db 7461 CTTTCATGAGCAGAGTCTTACCCAAAGAGAAAGCAAAACTTGAAGTACTGCTGGA 7520  
QY 608 uLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysAsnGlu 628  
Db 7521 AAACGACCCATGATAGATGACATCCCCATGATATGTTGAACTGCTAGCTAATAAACCA 7580



QY 628 nHisGluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgIleGlnSerIly 648  
 Db 7581 GCATGAGAGCAGCTTATGCTGACACTGATTGCTGACATCAACCGTATTCAATCCAA 7640  
 QY 648 sThrThrAlaAspAspCysValIleValAlaAlaLysAspGlySerAspTyrAlase 668  
 Db 7641 GACAACCTGCTGATGATGATGTTGTAATAGTAGCTGCCAAGAGTGGTTCAGATTATGCAATC 7700  
 QY 668 rSerValPheAspThrAsnSerGlnGlnLysSerLeuAlaSerGlnSerThrGlnLysGln 688  
 Db 7701 AAGTGTGTTTGACACTTAATCCCAACAGAGTCTCTTGSCATCCCAAGTACACAGAAGA 7760  
 QY 688 uLeuGlnGlyHisLeuAlaLeuThrThrGlnGlnSerProHisProGlnAsnPheGlnSe 708  
 Db 7761 GTTACAGGGTCAATTTGGCATTCACACAGAGTCTCCACATCTCCAGAACTTTTCAGTC 7820  
 QY 708 rThrGlnGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerSe 728  
 Db 7821 TACTCAGGAACACAGACACATTTGGGATGGAGAAATGGTCATCTATTGCTGCAAGCTC 7880  
 QY 728 rProLeuPheSerHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTrpGln 748  
 Db 7881 ACCACTATTTCACATCATCATCATCATATATGCTGAAGCACCACTCAACATTTGGGG 7940  
 QY 748 yArgLysAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSerPr 768  
 Db 7941 CGGTAAAGGACGCAAGAGCTAAGTGGGAGCAATTTAAGGCCACTACAGAAATTTCTCC 8000  
 QY 768 oAlaAlaThrCysGlyValaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerTh 788  
 Db 8001 AGCAGCAACATGGTGTCTCAATTTAGACTGGTATCCAGAGCATGTGACTTCACTTCTAC 8060  
 QY 788 rHisValMetGlySerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAs 808  
 Db 8061 TCATGTTCATGGGATCTTCAGCAATTTATCATCTCCCAACCAAGTAATTTGCGCCACTGA 8120  
 QY 808 pArgTyrAlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAl 828  
 Db 8121 CCCTATGCTGAAGAGCGGTAAACAGGTCCATGCAAGAAATTTTCCAGCACATATAGC 8180  
 QY 828 aThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValValLeuTyrPr 848  
 Db 8181 AACCATGGAAGCGAGTAAGTATGTGATCGGAGAAATGCTGGCAAGTAGTCTGTGTATCC 8240  
 QY 848 oLysGlnSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlase 868  
 Db 8241 TAAAGTAATCCATCCCTCGCGAGCATCTTCGAGATGATGGATCCATCAACATTAGCAAG 8300  
 QY 868 rPheProAsnTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerGl 888  
 Db 8301 CTTCCCACTATGGAACCTTAGCAGGAACCAAGATGGAGTCTCAACTTCATATTTCTCA 8360  
 QY 888 nTyrAlaHisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsnGl 908  
 Db 8361 GTATGCACATTAATCAGTACAAAGGATCAACACACATCATATATGGCAGTAACCTGAATGG 8420  
 QY 908 yLysIleProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgPr 928  
 Db 8421 AARGATTCATTCATTCAGAACTTATCAGCGCATCAGTCCATCATGATCGACAGACC 8480  
 QY 928 oLeuArgProHisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIleAlaAs 948  
 Db 8481 TTTACGCCACATCTAGAGTTGGTGTGCTTGGCTCTCTGTCAGAAAGGAATTCGAAA 8540  
 QY 948 nTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSerThrGlyIleTh 968  
 Db 8541 CTGGTTCGGAACATGTGGCACACAACTGGTTTATAGTTAGGAGTGTCCACAGATATAC 8600  
 QY 968 rSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPheSerAl 988  
 Db 8601 ATCCCATCAGATGAACAGAAAGCAATTTTGAAGCCCTGAATCTTGGAAATGTTTCAGC 8660  
 QY 988 aLysTrpAsnAlaLeuGlnLeuGlySerValSerSerValSerSerAlaAspPheLeuSerAlaAr 1008

Db 8661 AAAATGGAAATGCAATGCAATGGTTCCTGTTAGCTCCAGTCCAGATTTTATATCAGCGAG 8720  
 QY 1008 gAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHisProLeuAspAr 1028  
 Db 8721 GACACGATAGCTCATCTTGGACAGAGGACAGGTTAAATGGTTCAFCCTTCGATCG 8780  
 QY 1028 pPheValArgGlnAspIleCysIleThrAsnLysAsnProAlaAspPheThrThrIleSe 1048  
 Db 8781 GTTCTGTGAGACAGGATATCTGTATATACTAAACAAGAACCCAGCTGATTTTACTACAATCAG 8840  
 QY 1048 rAsnAspAsnGlnLysTyrMetAspTyrArg 1057  
 Db 8841 TAAAGTAACAGTATATGATGATACCGC 8868  
 RESULT 4  
 US-10-425-114-4042  
 ; Sequence 4042, Application US/10425114  
 ; Publication No. US2004003488A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 4042  
 ; LENGTH: 871  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: 700347436\_FLI  
 US-10-425-114-4042  
 Alignment Scores:  
 Pred. No.: 9,19e-37 Length: 871  
 Score: 491.50 Matches: 107  
 Percent Similarity: 65.79% Conservative: 18  
 Best Local Similarity: 56.32% Mismatches: 50  
 Query Match: 8.89% Indels: 15  
 DB: 13 Gaps: 6  
 US-09-721-114-2 (1-1057) x US-10-425-114-4042 (1-871)  
 QY 871 AsnTyrGlyThrSerSerArgAsnGlnMetGlu-----SerGlnLeuHisAsnSerGln 888  
 Db 2 AACTACCAAGAGCTAATAGCGCGCAGATGGAGCTTCAAACGCAAGCTCTCGGCTCGCAG 61  
 QY 889 TyrAla---HisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsn 907  
 Db 62 TACACACAGCATGAGCATTTACATATGCTCACCAGACATCATATGGAAGCCAA----- 115  
 QY 908 GlyLysIleProLeuThrPheGluAspLeuSer---ArgHisGlnLeuHisAspLeuHis 926  
 Db 116 -----CCGCTACACTGGAAGACTGTCTCGGGTCAATTCACGACAGACTTGGC 166  
 QY 927 ArgProLeuArgProHisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIle 946  
 Db 167 AGGCGCTTTACGCCCTCACCCCTCGTGTGGTGTGCTCGGTTCAATTCGTCAGCAGGAGATC 226  
 QY 947 AlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGlyValSerThrGly 966  
 Db 227 GGAACCTGCTCTGGGAACCTGCGGCGCGAGTCTGGTACGACTAGTGTATGTAAGGG 286  
 QY 967 IleThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPhe 986  
 Db 287 ACAAGCGCGCTGATGCCCAACAGAGCAGGAAACTACGAGACCTTCAGC----- 334



```

Db 1253 AAAAGGCAATCAGTCAAGCTCTATATTTCTAAAGAGAAACGCAAAACCCAGTCT 1312
Qy 461 -----ThrGlySerValHisHisThrValAlaHisPro 471
Db 1313 GAGTCTTTCTAATTATGACTCAGAAATTAGAAAAAGAGATAAAGACATGAGTAAATTTGGT 1372
Qy 472 AlaGlyAsnLeuSerAspLysValThrProThrAlaSer---ThrGlnHisAspAsp 490
Db 1373 GCTCCAGAACCCACCAAAAAGAAATCCAAATACAAAAGATTTTGACTCTTCTGAAGAT 1432
Qy 491 GluAsnAspThrGluAspGlyLeuAspThrAsnMetHisLysThr---AspValCysGln 509
Db 1433 GAGAAACACAGCAAAAAGGAATGATAATCAAGGGCACAAAATTTGAAGACCTCCAA 1492
Qy 510 HisValSerGluIleSerThrGlnArgCysSerSerLysGlyLysThrAlaGlyLeuSer 529
Db 1493 GAAGATCATCTGATGATGCTGAAGA---AAACAAGAGAGAGAGACTTCTCTTCAGCA 1549
Qy 530 LysGly-----Lys 532
Db 1550 GAAGCACAGTTGATAAGACACGACCATCATGGAATTTAAGAGATCGACTTCTTCAAGAAG 1609
Qy 533 ThrHisSerAlaAlaSerThrLysTyrrGlyGlyGlnSerThrArgAsnGlyGluAsnIle 552
Db 1610 CAGCAAGCAAGTGTCTCCACT-----GATGGTGTGATAGAGCTTCTTGGGAAGAGCAG 1663
Qy 553 HisValLeuSerAlaGluAspGlnCysGlnMetGluThrGlnAsnSerValLeuSerHis 572
Db 1664 AGTTTACTTCTTGGAGTTAGAAAGTCTCTGAACCTAAGAAAG-----AGCAAG 1717
Qy 573 SerAlaLysValSerProAlaGluHisAspIleGlnIleMetSerAspLeuHisGluGln 592
Db 1718 CATCTCAAAACCAAAACATGTAAAAAAGTACAGGATGGCTTATCTGATATTCGAGAGAA 1777
Qy 593 SerLeuProLysLysLys-----LysLysGlnLysLeuGlu 604
Db 1778 TTCCTAAGAAGACGACGAGGATGAACCTTCTGAAGATGATTAAGACGACGCAAAAAG 1837
Qy 605 ValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeu 624
Db 1838 GGAAGTGAAGAAAAAAGAAA-----CCTTCAGACTTTAAGAAAAAAGTA 1882
Qy 625 AlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSerAsp----- 641
Db 1883 ATTAAA-----ATGGAACAACAGTATGATATCTTCATCTGATGCGCATGAAAGTTACCT 1936
Qy 642 -----IleAsnArgIleGlnSerLysThrThr 650
Db 1937 GAGCGAGAGAAATTTGTCATTTCTTAAGCGCATATAACAAATTAAGAATGGAACAAT 1996
Qy 651 AlaAspAspAspCysValIleValAlaLysAspGlySerAspTyrAlaSerSerVal 670
Db 1997 -----GATGGAGAAAAGAAAAGTAAAAAATA 2023
Qy 671 PheAspThrAsnSerGlnGlnLysSerLeuAlaSerGlnSerThrLysGluLeuGln 690
Db 2024 AGAGATAAACTTTTAAAGAGAGATGAATATCTGATATGCTGAGAGTCAA--CAG 2081
Qy 691 GlyHisLeuAlaLeuThrThrGlnGlnSerProHisProGlnAsnPheGlnSerThrGln 710
Db 2082 GGA----- 2084
Qy 711 GluGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerSerProLeu 730
Db 2085 -----AAGAGNATGTTGTGACTCTT----- 2105
Qy 731 PheSerHisAspAspGlnTyrIleAlaGluAlaProThrGluHis--TrpGlyArgGly 750
Db 2106 -----CAGAGGATAAAAGAGTAAAGATGGAGCATATGTTAGAGAGA 2147
Qy 750 sAspAlaLysLysLeuThrTrpGluGlnPheLysAlaThrThrArg 765

```

```

Db 2148 AGAAAAGTGCAAGTTGCTTGGAAAGAGTTCAAGGAGAGACAAGA 2193
RESULT 6
US-10-252-157-188
; Sequence 188, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
; APPLICANT: Paris, Mary
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252,157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 188
; LENGTH: 11167
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030190640A1 261982.8
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 9131, 9138, 10638
; OTHER INFORMATION: a, t, c, g, or other
US-10-252-157-188
Alignment Scores:
Pred. No.: 8,13e-10 Length: 11167
Score: 223.00 Matches: 172
Percent Similarity: 34.31% Conservative: 108
Best Local Similarity: 21.08% Mismatches: 312
Query Match: 4.04% Indels: 225
DB: 15 Gaps: 33
US-09-721-114-2 (1-1057) x US-10-252-157-188 (1-11167)
Qy 62 ArgGlyTyrValAlaLeuLeuGlnLysAspProLysPheCysSerLeuSerArgIle 81
Db 1425 AAGGCTCACTTCATTCGATTCGGAAGAGACTTAAATTCGAGTTTCGAGCGATGCTGTGA 1484
Qy 82 PheHisAspGlnLysCysAspGluHisLysAlaSerSerSerProPheSerVal--- 100
Db 1485 ---AACAAAGAGAAAAATACCAAGAGCATAAAGTCATGATCTAGTCTTAAGTTGAAACAAA 1541
Qy 101 -----AlaLysPheArgArgTrpAspCysSerLysCysLeuAsp 113
Db 1542 GCACGAAAGGAGAAAAACCTTGTGCTTTGCGAAAGAGGATATTTCAAAGTCAGAAGCT 1601
Qy 114 LysLeu----- 115
Db 1602 AAACCTTCAAGAAAAACAGGTAGATAGTGCACATGCATCAGAAATGTTCCACAGAGGAA 1661
Qy 116 -----LysThrSerAspAsnGlyThrAlaPro 124
Db 1662 CAAAGAACAAATAAAGTACCGGTGGTGAACATCTGATAGAAAAGAGAACTT 1721
Qy 125 ArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIleThrPheValArg 144
Db 1722 CAATATGAACCTGCC-----AACACTCTTGAAGATTTAGACATGATATTGTGTCT 1772
Qy 145 SerThrPheValProAlaSerValGlySerGln-----LysValSerPro 159
Db 1773 -----GTTCTTCTCAGTCTCAGAGACATTTTGGAGAAATTTGAGACTGTGTATG 1823
Qy 160 SerThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGln 179
Db 1824 GAAGTTTCAGAGTTTCACTTGTATCATCAAGGGATGCGCAGCAGTGAACCTGAACAGAGATG 1883
Qy 180 GluGlyAsnAspSerLysCysAsnAlaProSerGlyLysAsn-----GlyAlaAlaGlu 197

```



```

; SOFTWARE: PERL Program
; SEQ ID NO 2759
; LENGTH: 279
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURES:
; NAME/KEY: misc_feature
; OTHER INFORMATION: incyte ID No. US20010051335A1 700346612H1
; NAME/KEY: unsure
; LOCATION: 28, 123, 150, 179, 191, 264
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-2759

Alignment Scores:
Pred. No.: 4,87e-12 Length: 279
Score: 222.50 Matches: 55
Percent Similarity: 64.50% Conservative: 7
Best Local Similarity: 57.25% Mismatches: 25
Query Match: 4.03% Indels: 10
DB: 9 Gaps: 4

US-09-721-114-2 (1-1057) x US-09-294-093B-2759 (1-279)
Qy 871 AsnTyrGlyThrSerSerArgAsnGlnMetGlu-----SerGlnLeuHisAsnSerGln 888
Db 2 AACTACCAAGAGCTTAATAGCGGCANATGGAGCTTCAACGCAAGCTCTCGGCTCGCAG 61

Qy 889 TyrAla---HisAsnGlnTyrLysGlySerThrSerThrSerThrSerThrSerThrSer 907
Db 62 TACACACAGCATGACCATTAACATGCGTCACCGACGAGCTCATATGGAAGCCAA----- 115

Qy 908 GlyLysIleProLeuThrPheGluAspLeuSer---ArgHisGlnLeuHisAspLeuHis 926
Db 116 -----CCGCTGAACTGGAAGACTTGTCGCGCTCATTCACCAAGACTTGGCG 166

Qy 927 ArgProLeuArgProHisProArgValGlyValGlyValGlyValGlyValGlyValGly 946
Db 167 AGGCTTTACGNCNCTCACCTCGANTCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 226

Qy 947 AlaIleTrpSerGluAsnCysGlyThrGlnSerGlyTyrLysLeuGly 962
Db 227 GCAAACTGGTCTGG-GAATGCGGCGCGCAGTCTGGGTANAGACTAGGT 273

RESULT 8
US-09-864-761-19041/c
; Sequence 19041, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecolica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19041
; LENGTH: 3953
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000511.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EST HUMAN HIT: AM867076.1, EVALUATE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUATE 5.00e-63
; OTHER INFORMATION: NT HIT: AE001609.1, EVALUATE 6.50e-01
US-09-864-761-19041

Alignment Scores:
Pred. No.: 3.59e-10 Length: 3953
Score: 220.00 Matches: 153
Percent Similarity: 35.32% Conservative: 125
Best Local Similarity: 19.44% Mismatches: 311
Query Match: 3.98% Indels: 198
DB: 9 Gaps: 38

US-09-721-114-2 (1-1057) x US-09-864-761-19041 (1-3953)
Qy 113 AspLysLeuLysThrSerAspAsnGlyThrAlaProArgThrLeuProAlaLysGlnAsn 132
Db 2346 GACAGGACTCTTGGCCAAATGAGAGACCCGCCATCTCTAGCAGAGCTCAGAGAAAT 2287

Qy 133 GlyThrSerAspGlyCysSerIleThrPheValArgSerThrPheValProAlaSerVal 152
Db 2286 GGACAA-----AGGACCCCATTTGCCAATGAGAGACC 2254

Qy 153 GlySerGlnLysValSerProSerThrGlnSerSerGlnGlyLysAsnAlaAspArgSer 172
Db 2253 ACATCATCTCTACAGAGCTTACAGAACACGAGAAAGGACTCCACTGGCCCAATGAGAAC 2194

Qy 173 ThrLeuProLysSerValGlnGlyAsnAspSerLysCysAsnAlaProSerGlyLys 192
Db 2193 ACCACACCA---TCCCGGCGAGAGCTTACAGAAATAGA-----GAA 2155

Qy 193 AsnGlyAlaAlaGluAlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGln 212
Db 2154 AGGACAGCCCAATGAGAGACACCAACCATCCCA-----GCAGGGCTTACAGAA 2107

Qy 213 AsnTyrAspValAlaAlaAsnValSerGluAsnThrSerValAspValGlyAlaLeu 232

```

2106	Db	AAATAGAAATATGACAGCCAAAC-----GAGAGAGACCACATATTCCAGCAGAGCCT	2055
233	Qy	ProGluValProGlnIleThrTTPHisIleGluValAsnGlyAlaaspGlnProPser	252
2055	Db	ACAGAAAATAGAGAAAGACACCAATGAGAGAGACACATCATCCCCAGCAGAGCCTACA	1996
253	Qy	-----ThrProLysLeuSerGluValValLeuLysArgenGluAspGlu	267
1995	Db	GAAAATGGACAAAGGACCCCATTTGGCAATGTGAAACACCATCATCTCCAGCAGAGCCT	1936
268	Qy	AsnGlyLysThrGluGluThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro---	286
1935	Db	ACAGAACACGGAGAGGACCCCACTGGCCAAATGAG---AACACCACTATCCCCAGCA	1879
287	Qy	AsnProMetSerGlyLysGlu-----	293
1878	Db	GAGCCTACAGAAAATAGAGAAAGGACAGCCCAATGAGAGAGACCACACCATTCACGAGAG	1819
294	Qy	-----ArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspPro-----	307
1818	Db	CCTACAGAAAATAGAGAAAGGACAGCCCAATGAG---AACACCACTATCCCCAGCAG	1762
308	Qy	-----LysProValSerGlyGlnLysCysGluGlnIleCysAsn	320
1761	Db	CCTACAGAAAATGGAGACAGAGCTCCATTGGCCAAATGAGAGAGACCACCATCTCTAGCA	1702
321	Qy	GluProCysGluGluValValLeuLysArgSerSerLysSerLysArgLysThrAspLys	340
1701	Db	GAGCCTACAGAAAATGGA-----AAAGAGCCCCATTGCCAATGAGAGAGACCATCA	1648
341	Qy	LysLeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAsp	360
1647	Db	TCCTCAGCAGAGCCTACAGAAACAGCGAGAAAG-----	1615
361	Qy	AlaLysLeuCysArgArgLysProLysLysValArgLeuLeuSerGluIleIleAsnAla	380
1614	Db	-----ACTCCACTGTCGCAATGAG	1597
381	Qy	AsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAla-----	396
1596	Db	AACACCACTATCCCGCAGCAGCCTACAGAAAATAGAGAAAGGACAGCCATGAGAAG	1537
397	Qy	-----AlaAspProCysGluAspAspArgSer-----Thr	406
1536	Db	ACCACCAATTCGCGCAGCAGCCTACAGAAAATAGAGAAAGCAGCCATGAGAAGACC	1477
407	Qy	IleProValProMetGluValSerMetAspIleProValSerAsnHis-----ThrVal	424
1476	Db	AGCAATTCGCGCAGAG-----CCTACAGAAAATAGAGAAATGAGATGAGAGCC	1432
425	Qy	GlyGluAspGlyLeuLysSerLysAsnLysThrLysArgLysTySerAspValVal	444
1431	Db	AATGAGAAACACCACTATCCCGCAGCAGAGCCTACAGAACT-----	1390
445	Qy	AspAspGlySerSerLeuMetAsnTTPLeuAsnGlyLysLysArgThrGlySerVal	464
1389	Db	GAAGAAATGACCCCATTTGGCCAAAT-----GAGAAGACCACTATCC-----	1348
465	Qy	HisHisThrValAlaHisProAlaGlyAsn-----LeuSerAsnLysLys	479
1347	Db	-----CCAGCAGAGCCTACAGAAAATGGAGAAAGGACCCCATTTACCAATGAGAAG	1297
480	Qy	ValThrProThrAlaSer-----ThrGlnHisAspAspGluAsnAspThrGluAsnGly	497
1296	Db	ACCACACCATCTCTCAGCAGAGCCTACAGAAACATGGAGAAAGGACCCCACTGGCCAAATGAG	1237
498	Qy	LeuAspThrAsnMetHisLysThrAspValCysGlnHisValSerGluIleSerThrGln	517
1236	Db	ATC---ACCACACCATCCCGCAGCAGAGCCTACAGAACTGGAGAAAGGATAGCCAAATGAG	1180
518	Qy	ArgCys-----SerSerLysGlyLysThrAlaGlyLeuSerLys	530
1179	Db	AAGGCCACACCATCCCCAGCAAAAGCCTTACAGAAACATGGAGAAACGACA---GTCAATGAG	1123

## RESULT 9

US-10-029-386-20602/c  
; Sequence 20602, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.

; TITLE OF INVENTION: HUMAN GEMCMB-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE OF INVENTION: AEMICA-X-2 EXPRESSION ANALYSIS TWO
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annotax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20602
; LENGTH: 3953
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AB023048.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUE 6.00e-63
; OTHER INFORMATION: EST HUMAN HIT: AW867076.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: AY026090.1, EVALUE 2.50e-01
US-10-029-386-20602

Alignment Scores:
Pred. No.: 3,59e-10 Length: 3953
Score: 220.00 Matches: 153
Percent Similarity: 35.32% Conservative: 125
Best Local Similarity: 19.44% Mismatches: 311
Query Match: 3.98% Indels: 198
DB: 15 Gaps: 38

US-09-721-114-2 (1-1057) x US-10-029-386-20602 (1-3953)

Qy	113	AspLysLeuLysThrSerAspAsnGlyThrAlaProArgThrLeuProAlaLysGlnAsn	132
Db	2346	GACAGGACTCTTTGGCCATGAGAACACACCGCCCTCTAGCAGAGCCTACAGAAAT	2287
Qy	133	GlyThrSerAspGlyCysSerIleThrPheValArgSerThrPheValProAlaSerVal	152
Db	2286	GGACAA-----AGGACCCCATTTGCCAATGAGAGACCC	2254
Qy	153	GlySerGlnLysValSerProSerThrGlnSerSerGlnGlyLysAsnAlaAspArgSer	172
Db	2253	ACATCATCTCCAGCAGAGCCTACAGAACACGAGAGAGGACTCCTCGCCCAATGAGAAC	2194
Qy	173	ThrLeuProLysSerValGlnGluGlyAsnAspSerLysCysAsnAlaProSerGlyLys	192
Db	2193	ACCACACCA---TCCCGCGCAGAGCCTACAGAAAATAGA-----GAA	2155
Qy	193	AsnGlyAlaAlaGluAlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGln	212
Db	2154	AGGACGCCAATGAGAACACACACCATCCCTCCCA-----GGAGGGCCCTACAGAA	2107
Qy	213	AsnTyAspValAlaAlaAsnValSerGluAspAsnThrSerValAspValGlyAlaLeu	232
Db	2106	AATAGAGAAATGACAGCCAC-----GAGAGAGCCACACTATTCCAGCAGAGCCT	2056
Qy	233	ProGluValProGlnIleThrTrpHisIleGluValAsnGlyAlaAspGlnProProSer	252
Db	2055	ACAGAAATAGAGAAAGGACGAGCAATGAGAGAACCCACATCATCTCCAGCAGAGCCTACA	1996
Qy	253	-----ThrProLysLeuSerGluValValLysLysArgAsnGluAspGlu	267
Db	1995	GAATAATGCAAGAGGCCCATTTGGCCATGAGAAACACCATCATCTCCAGCAGAGCCT	1936
Qy	268	AsnGlyLysThrGluGluThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro	286
Db	1935	ACAGAACACGAGAAAGGACCCCACTGGCCCAATGAG---AACACACACATCTCCCGACGA	1879
Qy	287	AspProMetSerGlyLysGlu-----	293
Db	1878	GAGCCTACAGAAATAGAGAAAGGACGAGCCCATGAGAGACCCACCATCTCCAGCAGAG	1819

Qy	294	-----ArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspPro-----	307
Db	1818	CCTACAGAAATAGAGAAAGGACGAGCAATGAG---AACACACACCATCTCCCGACGAG	1762
Qy	308	-----LysProValSerGlyGlnLysCysGluGlnIleCysAsn	320
Db	1761	CCTACAGAAATGAGACGAGACTCCATTGGCCCAATGAGAGACCCACCATCTCTAGCA	1702
Qy	321	GluProCysGluGluValValLeuLysArgSerSerLysSerLysArgLysThrAspLys	340
Db	1701	GAGCCTACAGAAATGGA-----AAAGGAGCCCATTTGCCAATGAGAGACCCACATCA	1648
Qy	341	LysLeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAsp	360
Db	1647	TCCTCAGCAGAGCCTACAGAACACGAGAAAG-----	1615
Qy	361	AlaLysLeuCysArgArgLysProLysLysValArgLeuLeuSerGluIleIleAsnAla	380
Db	1614	-----ACTCCACTGGCCCAATGAG	1597
Qy	381	AsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAla-----	396
Db	1596	ACACCATCATCTCCAGCAGAGCCTACAGAAATAGAGAAAGGACGAGCAATGAGAG	1537
Qy	397	-----AlaAspProCysGluAspAspArgSer-----	406
Db	1536	ACCACACAAATCCAGCAGAGCCTACAGAAATAGAGAAAGCAGACGCAATGAGAGACC	1477
Qy	407	IleProValProMetGluValSerMetAspIleProValSerAsnHis-----	424
Db	1476	ACACCATCTCCAGCAGAG-----CCTACAGAAATAGAGAAATGAGAGACGCC	1432
Qy	425	GlyGluAspGlyLeuLysSerSerLysAsnLysThrLysArgLysTySerAspValVal	444
Db	1431	AATGAGAACACACCATCTATCCCGCAGCAGAGCCTACAGAACAT-----	1390
Qy	445	AspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysArgThrGlySerVal	464
Db	1389	GAGAAATGAGCCCATTTGCCAAT-----GAGAGAGACCATCTATCTCC	1348
Qy	465	HisHisThrValAlaHisProAlaGlyAsn-----LeuSerAsnLysLys	479
Db	1347	-----CGAGCAGAGCCTACAGAAATGAGAAAGGACCCCATTTACCAATGAGAG	1297
Qy	480	ValThrProThrAlaSer-----ThrGlnHisAspAspGluAsnAspThrGluAsnGly	497
Db	1296	ACCACACCATCTCCAGCAGAGCCTACAGAACATGAGAAAGGACCCCACTGGCCCAATGAG	1237
Qy	498	LeuAspThrAsnMetHisLysThrAspValCysGlnHisValSerGluIleSerThrGln	517
Db	1236	ATC---ACCACACCATCTCCGAGCAGAGCCTACAGAACATGAGAAAGGAGATGCCAATGAG	1180
Qy	518	ArgCys-----SerSerLysGlyLysThrAlaGlyLeuSerLys	530
Db	1179	AAGGCCACACCATCTCCAGCAAGCCTACAGAACATGAGAAACGACA---GTCAATGAG	1123
Qy	531	GlyLysThrHisSerAlaAlaSerThrLysTyThrGlyGly-----	543
Db	1122	GACACACACCATCTCCAGCAGAGCCTACAGAAATGAGAAAGGACCCCACTGGCCCAAT	1063
Qy	544	-----GluSerThrArgAsnGlyGlnAsnIleHisValLeu	555
Db	1062	GAGAACACCAACATCTCCAGAGAGCTACAGAACATGAGAAAGGACGAGCAATGAG	1003
Qy	556	SerAlaGluAspGlnCysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLys	575
Db	1002	AAGACACACCATCTCCAGCAGAGCCTACAGAACATGAGAAAGGACCATCTAGCCCAAT	943
Qy	576	-----ValSerProAlaGluHisAspIleGlnIleMetSerAsp	588
Db	942	GAGAGACCATCTCTCCAGCAAGCCTACAGAACAGAA---GAATGAGCCCATCG	886
Qy	589	LeuHisGluGlnSerLeuProLysLysLysLysLys-----GlnLysLeuGlu	604

```

Db 885 GCATGAGAACCCACCATCTCCAGTAAAGCCTACAGAACATGAGAGAAAGACTACA 826
Qy 605 ValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeu 624
Db 825 TTGGCCCAATGAGAAGATCACTATCTCCCAAGAGGCGCT-----ACAGAACATGGA 775
Qy 625 AlaIysAsnGlnHisGluArgGlnLeuMetThrCluThrAspCysSerAspIleAsnArg 644
Db 774 GCAAAACTAGCTGGCCCAATGAGAAGATCACTATCTCCCAAGAGGCGCT-----ACAGAACATGGA 715
Qy 645 IleGlnSerLysThrThrAlaAspAspCysValIleValAlaAlaLysAspGlySer 664
Db 714 --GGAGAAAGGACCATCACTCCCAATGAGAAGATCACTATCTCCCAAGAGGCGCT-----ACAGAACATGGA 658
Qy 665 AspTyrAlaSerSerValPheAspThrAsn----- 674
Db 657 GAACATAGATAGGGCTCATCATGAGCAATGTGATCACACGCCCCCAGAGAGCTTATA 598
Qy 675 -----SerGlnGlnLysSerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlyHis 692
Db 597 AAACATGCAAAAGGACCATCTGGCCCATGAGAGATGACAAA----- 553
Qy 693 LeuAlaLeuThrThrGlnGlnSerProHisProGlnAsnPheGlnSerThrGlnGluGln 712
Db 552 -----GTCACAGAAAGTCCACAGAACACCCAGAAAGACCCAGTCAACACAGAGAAA 499
Qy 713 GlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerSerProLeuPheSer 732
Db 498 ACCACA-----AGAACCCAGAAAGCCTAAGCTATCTACAGAGAACCATATGCACC 445
Qy 733 HisHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTrpGlyArgLysAspAla 752
Db 444 AAAGGAGAAACACACACAGTCCACAGAAAGCCTACAGAAAGCCTGGGGAAC----- 394
Qy 753 LysLysLeuThrTrpGluGlnPheLeuAlaThrThrArgAsn-----SerProAlaAla 770
Db 393 ACCACATGACCTAGACCATTAAGCCCGGTAAGTCCACAGAAAGCCTGGGGAAC----- 334
Qy 771 ThrCysGlyAla-----GlnPheArgPro-----GlyIleGlnAlaVal 783
Db 333 ACAGCAGGATCAAAAGACTATAAAACCTTCAGTCAAGGTACAGAGGAGCAAAATCTCTC 274
Qy 784 AspLeuThrSerThrHisVal 790
Db 273 ACTACTACTCTCTCTCATCTA 253

RESULT 10
US-10-295-027-427
; Sequence 427, Application US/10295027
; Publication No. US2003023250A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
; CURRENT APPLICATION NUMBER: US/10295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15

; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 427
; TYPE: DNA
; LENGTH: 4383
; ORGANISM: Homo sapiens
; US-10-295-027-427

Alignment Scores:
Pred. No.: 4,15e-10 Length: 4383
Score: 220.00 Matches: 153
Percent Similarity: 35.32% Conservative: 125
Best Local Similarity: 19.44% Mismatches: 311
Query Match: 3.98% Indels: 198
DB: 16 Gaps: 38

US-09-721-114-2 (1-1057) x US-10-295-027-427 (1-4383)
Qy 113 AspLysLeuLysThrSerAspAsnGlyThrAlaProArgThrLeuProAlaLysGlnAsn 132
Db 1891 GACAGAGCTCTTTGGCCCAATGAGAAGACCGCCATCTCTAGCAGAGCTACAGAAAT 1950
Qy 133 GlyThrSerAspGlyCysSerIleThrPheValArgSerThrPheValProAlaSerVal 152
Db 1951 GGACAA-----AGACCCCATTTGCCCAATGAGAAGACCGCCATCTCTAGCAGAGCT 1983
Qy 153 GlySerGlnLysValSerProSerThrGlnSerSerGlnGlyLysAsnAlaAspArgSer 172
Db 1984 ACATCATCTCTCAGCAGAGCTTACAGACAGAGAAAGAGGACTCCACTGGCCCAATGAGA 2043
Qy 173 ThrLeuProLysSerValGlnGluLysAsnAspSerLysCysAsnAlaProSerGlyLys 192
Db 2044 ACCACACCA---TCCCGGCGCAGAGCTTACAGAAATAGA-----GAA 2082
Qy 193 AsnGlyAlaAlaGluAlaAsnThrAspSerProMetLysAspLeuGlnGlyProAlaGln 212
Db 2083 AGCAGAGCCCAATGAGAAGACACACCATCTCCCA-----GCAGGCGCTACAGAA 2130
Qy 213 AsnTyrAspValAlaAlaAsnValSerGluAspAsnThrSerValAspValGlyAlaLeu 232
Db 2131 AATAGAGAAATGACAGCCCAAC-----GAGAAAGACCACTATTTCCCGCAGAGCT 2181
Qy 233 ProGluValProGlnIleThrTrpHisIleGluValAsnGlyAlaAspGlnProSer 252
Db 2182 ACAGAAATAGAGAGAGGACAGCCCATGAGAGACCACTATCTCCCGCAGAGCTTAC 2241
Qy 253 -----ThrProLysLeuSerGluValValLeuLysArgAsnGluAspGlu 267
Db 2242 GAATAATGAGAAAGGACCCCATTTGCCCAATGAGAAGAAACCATCATCTCCCGCAGAGCT 2301
Qy 268 AsnGlyLysThrGluGlnThrLeuValAlaGluGlnCysAsnLeuThrLysAspPro 286
Db 2302 ACAGAACCGGAGAAAGGACCCCATCTGGCCCAATGAG-----AACACCACTATCTCCCGC 2358
Qy 287 AsnProMetSerGlyLysGlu----- 293
Db 2359 GAGCCTTACAGAAATAGAGAAAGGACAGCCCAATGAGAAGACCACTATCTCCCGCAGAG 2418

```



```
QY 294 -----ArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspPro----- 307
Db 2419 CCTACAGAAAATAGAGAGACGACCAATGAG---AACACCAACCATCCCGACGACAG 2475
QY 308 -----LysProValSerGlyGlnLysCysGluGlnLeuCysAsn 320
Db 2476 CCTACAGAAAATGAGACGACGACTCCATGGCCATGAGAGAGACCAACCATCTCTAGCA 2535
QY 321 GluProCysGluGluValLeuLysArgSerLysSerLysArgLysThrAspLys 340
Db 2536 GAGCTTACAGAAAATGGA-----AAAAGACCCCATTTGCCATGAGAGACCAATCA 2589
QY 341 LysLeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAsp 360
Db 2590 TCCTCAGCAGGCTTACGAGAACCGCGAAGG----- 2622
QY 361 AlaLysLeuCysArgArgLysProLysValArgLeuLeuSerGluIleLeuAsnAla 380
Db 2623 -----ACTCCACTGGCCCAATGAG 2640
QY 381 AsnGlnValGluAspSerArgSerAspGluValHisArgGluAsnAla----- 396
Db 2641 AACACCAATCATCCCGACGAGCGCTACAGAAAATAGAGAAAGGACAGCCCAATGAGAAG 2700
QY 397 -----AlaAspProCysGluAspArgSer-----Thr 406
Db 2701 ACCACCAATTCCTCAGCAGAGCGCTACAGAAAATAGAGAAAGCAGCCCAATGAGAGACC 2760
QY 407 IleProValProMetGluValSerMetAspIleProValSerAsnHis-----ThrVal 424
Db 2761 ACACCATTCCTCAGCAGAG-----CCTACAGAAAATAGAGAAATGAGAGAGCC 2805
QY 425 GlyGluAspGlyLeuLysSerLysAsnLysThrLysArgLysTyrSerAspValVal 444
Db 2806 AATGAGAGACCAACCATCTATCCCGACGAGCGCTACAGAAAT----- 2847
QY 445 AspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysLysArgThrGlySerVal 464
Db 2848 GAAGAAATGACCCCATTTGGCCAAAT-----GAGAGAGACCACTACTATCC--- 2889
QY 465 HisHisThrValAlaHisProAlaGlyAsn-----LeuSerAsnLysLys 479
Db 2890 -----CGACGAGCGCTTACAGAAAATGGAGAAAGGCCCCCATTTACCAATGAGAAG 2940
QY 480 ValThrProThrAlaSer-----ThrGlnHisAspAspGluAsnAspThrGluAsnGly 497
Db 2941 ACCACACCATCTCTCAGCAGAGCGCTACAGAAATGGAGAAAGGCCCCCACTGGCCCAATGAG 3000
QY 498 LeuAspThrAsnMetHisLysThrAspValCysGlnHisValSerGluIleSerThrGln 517
Db 3001 ATC---ACCACACCATCCCGACGAGCGCTTACAGAAATGGAGAAAGGATAGCCCAATGAG 3057
QY 518 ArgCys-----SerSerLysGlyLysThrAlaGlyLeuSerLys 530
Db 3058 AAGGCCACACCATCCCGACGAGCGCTTACAGAAATGGAGAAAGCGACA---GTCNAATGAG 3114
QY 531 GlyLysThrHisSerAlaAlaSerThrLysTyrGlyGly----- 543
Db 3115 GACACCAACCATCTCTCAGCAGAGCGCTTACAGAAAATGGAGAAAGGCCCACTGGCCCAAT 3174
QY 544 -----GluSerThrArgAsnGlyGlnAsnIleHisValLeu 555
Db 3175 GAGAAACCAACCAATCCCGACGAGCGCTTACAGAAATGGAGAAAGGACGAGCCCAATGAG 3234
QY 556 SerAlaGluAspGlnCysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLys 575
Db 3235 AAGACCAACCATCCCGACGAGCGCTTACAGAAATGGAGAAAGGACCACTAGCCCAAT 3294
QY 576 -----ValSerProAlaGluHisAspIleGlnIleMetSerAsp 588
Db 3295 GAGAAAGACCAATACCATCTCCAGCAAGCGCTTACAGAAACCGAAA---GAAATGACCCCATCG 3351
QY 589 LeuHisGluGlnSerLeuProLysLysLysLysLys-----GlnLysLeuGlu 604
```

```
Db 3352 GCCAATGAGAACACCAACCATCTCCCGAGTAAGCGCTTACAGAAATGGAGAAAGACTTACA 3411
QY 605 ValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeu 624
Db 3412 TTGGCCCAATGAGAGAGATCACACTATCCCGAGAGGCGCT-----ACAGAACATGGA 3462
QY 625 AlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArg 644
Db 3463 GCAAAAAATCACTCGCGCCCAATGAGAGAGATCACACCACTCCCTAGCAAAAGCCTCAGAACAT 3522
QY 645 IleGlnSerLysThrAlaAspAspCysValIleValAlaAlaLysAspGlySer 664
Db 3523 ---GGAGAAAGGACCAATCACCCCAATGACAGATCACCTCATCTCGACGAGAGTCTTACA 3579
QY 665 AspTyrAlaSerSerValPheAspThrAsn----- 674
Db 3580 GAACATAGAGATAGGCGCTATCATCAGCCCAATGTGATCACACAGCCCGACGAGCCCTATA 3639
QY 675 -----SerGlnGlnLysSerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlyHis 692
Db 3640 AACATGCAAAAGGACCACTTGGCCCATGAGAGATGACACAA----- 3684
QY 693 LeuAlaLeuThrThrGlnGluSerProHisProGlnAsnPheGlnSerThrGlnGluGln 712
Db 3685 ---GTCACAGAAAAGTCCACAGAACACCCAGAAAAGACCACTCAACCCACAGAGAAA 3738
QY 713 GlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSerSerProLeuPheSer 732
Db 3739 ACCACA-----AGAACCCCGAGAAAGCGCTACTATCTCAGAGAGAACCATATGCACC 3792
QY 733 HisHisAspAspGlnTyrIleAlaGluAlaProThrGluHisTrpGlyArgLysAspAla 752
Db 3793 AAGGGGAAAAACACACCACTCCCGAGAAAGCGCTACAGAAAACCTGGGGGAAAC----- 3843
QY 753 LysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsn-----SerProAlaAla 770
Db 3844 ACCACACTCACCCTGAGACCACTAAAGGCCCGAGTAAGTCCACAGAAACCCAGAGAAA 3903
QY 771 ThrCysGlyAla-----GlnPheArgPro-----GlyIleGlnAlaVal 783
Db 3904 ACAGCAGCAGTCAAAAGACTATTAACCTTCAGTCAAGGTCACAGAGACAAATCTCTC 3963
QY 784 AspLeuThrSerThrHisVal 790
Db 3964 ACTACTACCTCTCTCTCATCTA 3984

RESULT 11
US-10-120-988-286
; Sequence 286, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Felyan
; APPLICANT: Dmanac, Raduje T.
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and
; FILE REFERENCE: 802CON
; CURRENT APPLICATION NUMBER: US/10/120,988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: pt FL_gense Version 2.0
; SEQ ID NO 286
; LENGTH: 6745
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
```

QY	297	ValAlaGluGlnCysAsnLeuThrLysAspProLysProValSerGlyGlnLysCysGlu	316
DB	3850	ACATTACATAAAAGTTAAG---ACGACTAGTACGTTTCTGTTCTGGTATGAGTAAAT	3906
QY	317	GlnIleCysAsnGluPro-CysGluGluValValLeuLysArgSerSerLysLys	335
DB	3907	GTAAATGTCGGAGTGTCTCAATATATTACACTCTCCAGGAAACCCAGCAAAAA	3966
QY	335	-----	335
DB	3967	FTCTGTAACTCCTTCAACAGTATACACAAATATAATTACTTATAGAAATCACTCAA	4026
QY	336	-----ArgLysThrAspLysLeuMet	343
DB	4027	GTGGAGACTGAACATTTCTTAACGCTTTAGAAAGACAAACAGAAATATTTCACACGA	4086
QY	344	-LysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAspAlaLysLe	363
DB	4087	GAGCAGTCCAGAACACCTTCATGTGAAATCTAAAGATGT-----CCGTCAACTCT	4137
QY	363	uCysArgLysProLysValArgLeuLeuSerGluIleLeuAsnAlaAsnGlnVa	383
DB	4138	GATCAGACCTCACCACCTGAAATATGACTGCTCCGATTATCA-AAATAGGGGGCCCT	4196
QY	383	lGluAspSerArgSerAspGluValHisArgGluAsnAlaAlaAspProCysGluAspAs	403
DB	4197	AGCTCCTACATTACAGGAATGCTTCTGTTGAGGCGCT-----	4236
QY	403	parSerThrIleProValProMetGluValSerMetAspIleProValSerAsnHisTh	423
DB	4237	-----GTTTCTTCTTCTGAGAGGAATCTAAAGCTAGAGATTTTTCAGATAA	4286
QY	423	rValGlyGluAspGlyLeuLysSerSerLysAsnLysThrLysArgLysTySerAspVa	443
DB	4287	TTTAGTAAACACCTCTAGGTGATTTCAGAAACCAAGAGGAAGA-----	4332
QY	443	lValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysArgThrGlySe	463
DB	4333	-----GCCAAAGATTCCAAAGTGAAC	4355
QY	463	rValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysValThrProTh	483
DB	4356	CCTGCATACCTTCTTCTGAGAGAAAAAATGTATCCGAGAAAAA---TCTGAAAA	4412
QY	483	rAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeuAspThrAsnMethi	503
DB	4413	TTGTCAACCAATCCATTATTCAGTAACAGTGTCTCTCTTCTTCCAGCT---CTTTC	4469
QY	503	sLysThrAspValCysGlnHisValSerGluIleSerThrGlnArgCysSerSerLysG	523
DB	4470	AGAGTATATATGGAAATTCCTCAACTAGAGAAATCTCTGGAGTGTACAGGGAGTGG	4529
QY	523	YLysThr-----AlaGlyLeuSerLysGlyLysThrHisSerAlaLase	538
DB	4530	TAGAGCCATTCCATTCTTACGGAAGTGTCCCAAGAGTGTCCCAAGAGATCACACATCCACG	4589
QY	538	rThrLysTyGly-----GlyGluSerThrArgAsnGlyGlnAsnIleHisValLeuSe	556
DB	4590	TGTAGTGTATGCTCCAGTGGATCAGAGCTAGGAGGACAGAGG-----GA	4637
QY	556	rAlaGluAspGlnCysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLys--	575
DB	4638	CATTGGACCAACTGCCAAAAAATGACT---AATAAAACACTTCTCTCACTCAGAGACTCA	4694
QY	576	-----ValSerProAlaGluHisAspIleGlnIleMetSerAspLeuHisGlu--	591
DB	4695	AGTCTTTGGCTTCTTCTCCAGCATTGCTAAACTACAGCTTGTGGAGGACTCAGTCAGA	4754
QY	592	-----GlnSerLeuProLysLys-----LysLysLysGlnLy	602
DB	4755	TGAACCAAACTTAGAGAGTCTGAGTCTGAAACCAAGAGAAATTAACCTCAAGAGTCA	4814
QY	602	sLeuGluValThrArgGluLysGlnThrMetIleAspAspIleProMetAspIleValG	622

QY	83	HisAspGln-----LysLysCysAspGluHisLysAlaSerSerSerProPheSer	99
DB	3034	CATGATGAAGGTGTGATGTCATTAAGCCATTACACACTCTCTTTTAGGAATGAAGAGGA	3093
QY	100	ValAlaLysPheArgArgTrpAspCysSerLysCysLeuAspLysLeuLysThrSerAsp	119
DB	3094	AAAGGAAATAAGGCATCAT-----ATATCTGTATTGAAAGTTTAAGCAAAACAGAA	3147
QY	120	AsnGlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSer	139
DB	3148	AGTATATACAGTACCC-----ACCAGTGTATCAGAGGC	3180
QY	140	IleThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysValSer---	158
DB	3181	CTC-----ATTGAGCAATATCAAGCAATTCCTCAAGTTCTTGAA	3219
QY	159	ProSerThrGlnSerSer-----GlnGlyLys	167
DB	3220	CTTGACACAATTTATTGTACCTTCCCAAGAAATCAAGCAGTTTCTCATACATGCGCAGG	3279
QY	168	AsnAlaAspArgSerThrLeuProLysSerValGlnGluGly-----	181
DB	3280	CAGTCAGGAAGTAAATAAGCTGCTTCATTAGGAATGGGCCACCTCCCTTCCAAATC	3339
QY	182	-----AsnAspSerLysCysAsnAla	188
DB	3340	AAAAATAATGTGAAGATGCAATGGGAATATATATGTTAAACAAATTTAGTCCCGTTCT	3399
QY	189	ProSerGlyLysAsnGlyAlaAlaGluAlaAsnThrAspSerProMetLysAspLeuGln	208
DB	3400	CCTGAGTCAGGAATGAATGTTCCTCAAGTCTTTCAGACTCAGCC-----CTGGAA	3450
QY	209	GlyProAlaGlnAsnTyAspValAlaAlaAsnValSerGluAspAsnThrSerValAsp	228
DB	3451	GCACCTGAAGCCACAGAGAGAATGCAAAATGTAAAAAGCAGTGGATCTACTTCCGTTAGA	3510
QY	229	ValGlyAlaLeuPro-----GluValProGlnIleThrTrpHis	241
DB	3511	AAAGGACCACTTCCATTCTCATCAAGGCTATGTCATGCTCCCTCAGGGGAGCCACAT	3570
QY	242	IleGlu-----ValAsnGlyAlaAspGlnProPro	251
DB	3571	GCCTCACTGGNAGAGAGGAGAAAAAGCCATTGACCTCAGGCATGGATGCTTCTGAG	3630
QY	252	SerThrProLysLeuSerGluValValLeuLysArgAsnGluAspGluAsnGlyLysThr	271
DB	3631	CTAACACCAAGGGCTTGGGAGAGAATCATTAGCCCTGTGGAAAGTGAATCATCT-	3684
QY	272	GluGluThrLeuValAlaGluGlnCysAsnLeuThrLys-----	284
DB	3685	-----GTTAGAGATTGTTCTTTTACCAAAAGACACCAACCAAGGAAAC	3729
QY	284	-----	284
DB	3730	TTCCAAGAAATACACTGAGAAAGAGGTAATAATGGCTGCTCCAGGAGAGTGTATTGCT	3789
QY	285	-----AspProAsnPro-----MetSerGlyLysGluArgAspGln	296
DB	3790	CTTTCAAATGAAGACCTTTACCTTTGCTCAGACTTGTCCAGAAAGAAAGACGTTGGAAA	3849

QY	9.51e-10	Length:	6745
Score:	219.00	Matches:	213
Percent Similarity:	33.71%	Conservative:	172
Best Local Similarity:	18.65%	Mismatches:	410
Query Match:	3.96%	Indels:	348
DB:	16	Gaps:	53

US-09-721-114-2 (1-1057) x US-10-120-988-286 (1-6745)	
---	--

Db 4815 GGCAAAATATGACAGAGCAGGAGG 4839  
Qy 622 uLeuLeuAlaLysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSerAsp11 642  
Db 4840 -----GCTGAAGATGAATGAGAGAGTCACTTGGGATCACTTCTCTGAAGG 4892  
Qy 642 eAsnArgGlnSerLysThrAlaAspAspCysValIleValAlaLysAs 662  
Db 4893 AAACAAA-----AATAAAACCAACTTGGATGAC-----CTAGTAAAGGGGAGAA 4937  
Qy 662 pGlySerAspTyr-----AlaSerSerValPheAspTh 673  
Db 4938 TAGATCTTCAGTTAAACACAGATTGGCGCATGTCTTAAGCCAGCAGAAAATCCAGC 4997  
Qy 673 rAsn-----SerGlnGlnLysSerLeuAlaSer-----GlnSerThrGlnLysG1 688  
Db 4998 TAAAGATGTAAGCCCAAGACAGATGAGTACTATCTCTCCCAAGTGAAGCAGATC 5057  
Qy 688 uLeuGlnGlyHisLeuAlaLeuThrThrGlnGlu-----SerProHi 702  
Db 5058 TGGCTTTGACCAATTTATCTCTGACAGTGGAGTGCACCCACTGTCTCCCTGAGCCTAC 5117  
Qy 702 sProGlnAsnPhenGlnSerThrGlnGlnGln-----ThrHisLeuAr 717  
Db 5118 TCCAAATCTGCAGAGTCCATTGGCAAGCAGGTTGAGTGAGAACGATGTGAA 5177  
Qy 717 gMetGluGlu-----MetVa 722  
Db 5178 GAAATCCGAGAACCTTCTCCCATTAAGTACTACCAACAGAGAACCTTCTACACAGT 5237  
Qy 722 lThrIleAlaAlaSerSerProLeuPheSerHisAspAspGlnTyr-----IleAl 740  
Db 5238 CAGCAACAGAGTCTAACGATCTTCAACGACATTCAGAGTGAAGTTTAAACGCTC 5297  
Qy 740 aGluAlaProThrGluHisTrpGlyArgLysAspAlaLysLysLeuThrTrpGluGlnPh 760  
Db 5298 AGATCACCATCAAGCATCAGATTCTAAAGAC----- 5331  
Qy 760 eLysAlaThrThrArgAsnSerProAlaAlaThrCysGlyAlaGlnPheArgProGly11 780  
Db 5332 ----GTACAGCAGCTCAGATTAGTAAGAGAAATCAGGAGCCCA----- 5373  
Qy 780 eGlnAlaValAspLeuThrSerThrHisValMetGlySerSerSerAsnTyrAlaSerAr 800  
Db 5374 -TCACCCATCATCTCACCAGCTCAGGAGCAGAGATTCTCTGACAACTCAGAGAGGCT 5432  
Qy 800 gGlnProValIleAlaProLeuAspArgTyrAlaGluArgAlaValAsnGlnValHisAl 820  
Db 5433 GAGCCCTCTCTTT---CCACTGGAGCT---GCACAGAAA----- 5466  
Qy 820 aArgAsnPhenProSerThrIleAlaThrMetGluAlaSerLysLysLeuCysAspArgAs 840  
Db 5467 -----TCTAGAGTAAGCAGTCCACTGCGCAGTTTCTGAGCAACAAAGAG 5513  
Qy 840 nAlaGlyGlnValValLeuTyrProLysGluSerMetProAlaThrHisLeuLeuArgMe 860  
Db 5514 TGCTTCATCTCTG-----GAGTGGGAACCTGAGCCACACCTCTATCTG-- 5556  
Qy 860 tMetAspProSerThrLeuAlaSerPheProAsnTyrGlyThrSerSerArgAsnGlnMe 880  
Db 5557 -----TCRAAGAGTTTAAAGACATTAATGTTCTATGCGCATCTACTACGAAAAGC-- 5607  
Qy 880 tGluSerGlnLeuHisAsnSerGlnTyrAlaHisAsnGlnTyrLysGlySerThrSerTh 900  
Db 5608 -----CATCTCCAAAAGTCCAGGAGCGGCATTTTCTGAAGACACTTCTAT 5654  
Qy 900 rSerTyrGlySerAsnLeuAsnGlyLysIleProLeuThrPheGluAspLeuSerArgHi 920  
Db 5655 T-----GACAATGCCCTGAGTCGACTGACCTTGGGAATGAATTTCTGTCAACAA 5705  
Qy 920 sGlnLeuHisAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlySe 940

Db 5706 TGGGTACAGTCGAAGATTGAGATCTTTTCTGAACCTCCCTCTGTGAT-----GGAAA 5759  
Qy 940 rLeuLeuGlnLysGluIleAlaAsnTrpSerGluAsnCysGlyThrGlnSerGlyTyrLy 960  
Db 5760 T-----GAAAGTTGGCTTATCGACGGGACAAACAGAGTCCCGAG 5801  
Qy 960 sLeuGlyValSerThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAl 980  
Db 5802 GTCTGCAATATCTATA-----TACAGACC 5825  
Qy 980 aLeuAsnSerGlyMetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSe 1000  
Db 5826 TATCGACTATGGATCTTTGGGAAAGAACACACAGTAGCTTCTTAGAAAAATGTAAGAG 5885  
Qy 1000 rSerAlaAspPheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysG1 1020  
Db 5886 GTCA-----CTTACACAGGAGATATGGAACCAAGTTTCTTAAAGAACCTCG 5936  
Qy 1020 yLysMetValHisProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAs 1040  
Db 5937 C-----TTCCTAAAAGATGATTG-----AGGAA 5960  
Qy 1040 nPro 1041  
Db 5961 CCT 5964  
RESULT 12  
US-09-945-917-51  
; Sequence 51, Application US/09945917  
; Publication No. US20030042361A1  
; GENERAL INFORMATION:  
; APPLICANT: Bogaert, Thierri  
; APPLICANT: Vandekerckhove, Joel  
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH  
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND  
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR  
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR  
; FILE REFERENCE: P/14-1  
; CURRENT APPLICATION NUMBER: US/09/945,917  
; CURRENT FILING DATE: 1998-09-21  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 51  
; LENGTH: 4584  
; TYPE: DNA  
; ORGANISM: Caenorhabditis elegans  
US-09-945-917-51  
Alignment Scores:  
Pred. No.: 1,63e-09 Length: 4584  
Score: 214.00 Matches: 243  
Percent Similarity: 34.68% Conservative: 170  
Best Local Similarity: 20.40% Mismatches: 414  
Query Match: 3.87% Indels: 365  
DB: 10 Gaps: 56  
US-09-721-114-2 (1-1057) x US-09-945-917-51 (1-4584)  
Qy 14 ValGlyThrAsnCysMet-----LeuAlaArgGlyGlyThrGlyAla 27  
Db 247 CTGGTCTCGACTGCTCGAAACTCACCAAAACCGATATCGACAGCGGAAACTTGGGTGCA 306  
Qy 28 ValAlaProValLeuGluLeuThrAlaThrProArgGlnAspAlaAlaGluAlaGly 47  
Db 307 GTTCTCCAGCTGCTCTTCTGCTCTCCACTTACAGCAG----- 345  
Qy 48 ValAspGluProAlaGlnHisGlnCysGluHisPheSerIleArgGlyTyrValAlaLeu 67  
Db 345 ----- 345  
Qy 68 LeuGlnLysLysAspProLysPheCysSerLeuSerArgIlePheHisAspGlnLys 87  
Db 346 -----AAGCTTCGCAACTGAAAAAAGATCAGAGAAA 378





184 QY -----SerLysCysAsnAlaProSerGlyLysAsnGlyAlaAlaGluAla 198  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
199 AsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAlaAla 218  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
200 AsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAlaAla 218  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
219 AsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIle 238  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
220 AsnValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIle 238  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
239 ThrTrpHisIleGluValAsnGlyAlaAspGlnProSerThrProLysLeuSerGlu 258  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
259 ValVal-----LeuLysArgAsnGluAspGlnAsnGlyLysThrGluGluThrLeu 275  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
260 ValVal-----LeuLysArgAsnGluAspGlnAsnGlyLysThrGluGluThrLeu 275  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
276 ValAlaGluGlnCysAsnLeuThrLysAspPro-----AsnProMetSer 290  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
298 CTGAATTAAGATTATTCAGTAGCAAAACCCATCTTCCTCATCGAATAGCCCAACACCT 987  
QY GlyLysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLysProVal 310  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
988 ACAGAAAGCGCGCGCGCTCTCAACACAAACCTTTGCGAAATCGCTGCCCGCAGTG 1047  
QY -----SerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluValValLeuLys 329  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
1048 AAAAGTGGCTGAG-----CCGCGCACCGAGTAACTGAGCTGGGAAGT 1086  
QY ArgSerSerLysSerLys-----ArgLysThrAspLysLys 341  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
1087 GCCACGCTATGTCGAGCTTTGTACGCCAAAGATTTCCTACCGTAAACGAGCGGCCCA 1146  
QY LeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAspAla 361  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
1147 ATCATATCTCAACAA-----GACTCG 1167  
QY LysLeuCys-ArgArgLysProLysLys-ValArgLeu-----LeuSerGluIleAla 379  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
1168 AAACGATGCTCAAGACGATGAGAAAGAGTCCGGATACCGTGGATTCACACACAGCTC- 1226  
QY snAlaAsnGlnValGluAspSerArg-SerAspGluValHisArgGluAsnAlaAlaAsp 398  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
1227 --GCCAAGCTCATCATCGACGGAAGTTCCCTAAGCATGATTCACATCTTCCAGAGT 1284  
QY ProCysGluAspAspArgSerThrIleProValProMetGluValSerMetAspIle---- 417  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
1285 TCAACGCTCAGACGAAAGTCTCGTCAATCAGACGATCTTACTCTTAAACGCTCCATCGTG 1344  
QY 418 -----ProValSerAsnHisThrValGlyGluAsp 427  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
1345 ACAGCTATCAGACGCGATAGCGGCAACCGGTTCTCCAATATTATCAACAGCTT 1404  
QY 428 -----GlyLeuLysSerSerLysAsnLysThrLysArgLysTyr 440  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
1405 GTTCAGGAAAAACCAACACTGGCAGTCAAGAGGAGTGAAGGACACAGGAAAAAAGATCCA 1464  
QY 441 SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysArg 460  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
1465 CTCCAGCTGTCGCGACGAGTACCC-----CCGCCACA 1500  
QY 461 ThrGlySerValHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal 480  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
1501 ATCGGAGTTGTTAGTCAATATTAGGCACAT-----AAGAGTTG 1539  
QY 481 Thr-----ProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeu 498  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 729  
1540 ACAATGACCCCTGTATCTGAAACACAGAACTCTGAAAGCTCCCAATCAATAGGCATC 1599

499 AspThrAsnMetHisLysThrAspVal--CysGlnHisValSerGluLysSerThrGln- 517  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 1647  
518 -----ArgCysSerSerLysGlyLysThrA 526  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 1647  
1648 CTTAAATGACTTCAATCCGACCAACACACACACACACACACACACACACACACACACAC 1703  
QY 526 laGlyLeuSerLysGlyLysThrHisSerAlaAspSerThrLysTyrGlyGlyGlySer 546  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 1703  
1704 -----AAAAATCATCGCTGTCAAG--TCGTTTGGATATGACAGT 1744  
QY 546 hrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrG 566  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 1744  
1745 CG-----TCGCGCTG 1756  
QY 566 LuAsnSerValLeuSerHis--SerAlaLysValSerPro-----AlaG 580  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 1756  
1757 AAGATCCATTTGGCTCATGCTGGCTCAGTGTGCTCGGACCTCGCGCAACAACTTCTGTA 1816  
QY 580 LuHisAspIleGlnIleMetSerAspLeuHisGluGlnSerLeuProLysLysLysLysL 600  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 1816  
1817 ATCATTTGCTG-----GAGAGAGGATGGGAAGATAAGACAT 1855  
QY 600 ysGlnLysLeuGluValThr-----ArgGlu 609  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 1855  
1856 CAGATCCAGCGCTACACTCTGACGCGGTGTCGATGTGCGCAAAATGAGGAGA 1915  
QY 609 ysGlnThrMetIleAspAspIleProMetAspIleValGluLeuLeuAlaLysAsnGlnH 629  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 1915  
1916 AGCTGAAAGATACATGACATGACTGT-----CGAGCACAGACGGCT 1960  
QY 629 is-----GluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgI 645  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 1960  
1961 ATCTGACAACTTCGAGACAGTTCCTCTGCTGCTGGAATATCCGATATCAACAGAGC 2020  
QY 645 leGlnSerLysThrThrAlaAsp-----AspAspCysValIleValAlaAlaLysA 662  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 2020  
2021 TCGACGACATATCCAGGACAGTTCGCGAGTAGACATGCGCAACAGCTGCTCCAAAC 2080  
QY 662 spGlySerAspTyrAlaSerSerVal---PheAspThrAsnSerGlnGlnLys----- 678  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 2080  
2081 AT---AGGCACTATTCCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2137  
QY 679 -----SerLeuAlaSerGlnSerThrGlnLysGluLeuGlnGlyHisLeuAlaLeuThr 697  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 2137  
2138 TCCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2184  
QY 697 hrGlnGluSerProHisProGlnAsnPheGlnSerThrGlnGluGlnGlnGlnHisLeuA 717  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 2184  
2185 --CAGGAGATGTGTACAAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2235  
QY 717 rgMetGluGluMetValThrIleAlaAlaSerSerProLeuPheSerHisHis----- 734  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 2235  
2236 -----GCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2272  
QY 735 --AspAspGlnTyrIleAlaGluAlaProThrGluHisTyrGlyArgLysAspAlaLysL 754  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 2272  
2273 GATCCCGGAGTACTCATCTCTTCTCA-----CACTATCA---GTGCTGCTGCTGCTGCTG 2323  
QY 754 ysLeuThrTrpGluGlnPheLysAlaThrArg-----AsnSerProAlaAlaThrC 772  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 2323  
2324 AGGACACATGTCTATGACATCAAGACTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2383  
QY 772 ysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisValMetG 792  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 2383  
2384 ATTCAGGCAATTT-----CATTCCTGAT---CGTAAATGCCACCTTCAAG 2428  
QY 792 LysSerSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAspArgTyrAlaG 812  
Db TCGACGATATCCATCATCTCGAAGAGTTAGAAATCATCATCAACGTACAGCTCTATTTCG 2428  
2429 AGTTCACATCCACGAGACAGAGTGGGCTCTCTGAGCGCG----- 2472  
QY 812 LuArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGluA 832



Qy	276	ValalaGluGlnCysAsnLeuThrLysAspPro-----AsnProMetSer	290
Db	928	CTGAATATAAAGATTATTTCAGTAGCAAAAACCCATCTTCCTCATCGAATAGCCCAACACCT	987
Qy	291	GlyLysGluArgAspGlnValAlaGluGlnCysAsnLeuThrLysAspProLysProVal	310
Db	988	ACGAGAAGGCGCGCGGCTCTCAACACAAACTTTGTCGAAATCGCTGCCCGCAGTG	1047
Qy	311	---SerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluValValLeuLys	329
Db	1048	AAAAGTGGCCTGAAG-----CCGCGCAGCACTAAGCTGGGAAGT	1086
Qy	330	ArgSerSerLysSerLys-----ArgLysThrAspLysLys	341
Db	1087	GCCACGCTGATGCGAAGCTTTGACGCCAAAAGTTTCCTACCGTAAACCGACGCGCCCA	1146
Qy	342	LeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAspAla	361
Db	1147	ATCATATCTCAACNA-----GACTCG	1167
Qy	362	LysLeuCys-ArgArgLysProLysLys-ValArgLeu-----LeuSerGluIleIleA	379
Db	1168	AAACGATGCTCAAGACAGTGAAGAGAGTTCGGATACGCTGGATTCAACACAGCAGCTC-	1226
Qy	379	snAlaAsnGlnValGluAspSerArg-SerAspGlnValHisArgGluAsnAlaAlaAsp	398
Db	1227	--GCCAAGCTCATCTCAGCGAAGTTCCCTAAGCATGATTTCCACATCTTCCAGAGT	1284
Qy	399	ProCysGluAspArgSerThrIleProValProMetGluValSerMetAspIle---	417
Db	1285	TCAACGTCAGACGAAAAGTCTCCGTCATCAGACGATCTTACTCTTAACGCGCTCCATCGTG	1344
Qy	418	-----ProValSerAsnHisThrValGlyGluAsp	427
Db	1345	ACAGCTATACACAGCGCATAGCGCAACACCGGTTTCTCCAAATATATCAACAACGCT	1404
Qy	428	-----GlyLeuLysSerSerLysAsnLysThrLysArgLysTyr	440
Db	1405	GTTGAGGAAAACCAACACTGGCAGTGAAGGAGTGAAGACGACAGCGAAAAAGATCCA	1464
Qy	441	SerAspValValAspAspGlySerSerLeuMetAsnTrpLeuAsnGlyLysLysLysArg	460
Db	1465	CCTCCAGCTGTTCGCCACGCTGACACC-----CAGCCACA	1500
Qy	461	ThrGlySerValHisHisThrValAlaHisProAlaGlyAsnLeuSerAsnLysLysVal	480
Db	1501	ATCGGAGTTGTAGTCCAAATATGGCAGAT-----AGAAGTTG	1539
Qy	481	Thr-----ProThrAlaSerThrGlnHisAspAspGluAsnAspThrGluAsnGlyLeu	498
Db	1540	ACAAATGACCCCGTGATATCTGAAAAACACAGAACTCCAAAGGCTCCAAATCAATGAGCATC	1599
Qy	499	AspThrAsnMetHisLysThrAspVal--CysGlnHisValSerGluIleSerThrGln-	517
Db	1600	GACACG-----ACGACGTTCCACCGCTTCCACTCTTAAATCAGTTGTTCCTCA	1647
Qy	518	-----ArgCysSerSerLysGlyLysThrA	526
Db	1648	CTTAAATGACTTCAATCCGACACCAACCAAGTATGTTCTCTTAAACAAGG----	1703
Qy	526	laGlyLeuSerLysGlyLysThrHisSerAlaAlaSerThrLysTyrGlyGlyLysSerT	546
Db	1704	-----AAAAATCACATCCGCTGTCAAG--TCGTTGGATATGAGCAGT	1744
Qy	546	hrArgAsnGlyGlnAsnIleHisValLeuSerAlaGluAspGlnCysGlnMetGluThrG	566
Db	1745	CG-----TCCGCGTCTG	1756
Qy	566	luAsnSerValLeuSerHis---SerAlaLysValSerPro-----AlaG	580
Db	1757	AAGACTCCATTTGGCTCATCGTCGCTGAGTGACTCGCGCAGCAAAAACCTTCTGSTA	1816

Qy	580	luHisAspIleGlnIleMetSerAspLeuHisGlnGlnSerLeuProLysLeuLysLysL 600
Db	1817	ATCATTTGCTG-:::--GAGAGAAGTATGGGAAGAATAACACAT 1855
Qy	600	ysGlnLysLeuGluValThr-:::--ArgGluL 609
Db	1856	CAGAAATCCAGCGGCTACACCTCTGACGCGGTGTTCGATGTGCGCAAAATGAGGAGA 1915
Qy	609	ysGlnThrMetIleAspAspIleProMetAspIleValGlnLeuLeuAlaLysAsnGlnH 629
Db	1916	AGCTGAAGAATAACGATGACATCACTCGT-:::--CGAGCAGACAACGGCT 1960
Qy	629	is-:::--GluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgI 645
Db	1961	ATCCTGACAACCTCGAAGACAGTTCCTCTGCTGCTGGAATATCGATAACAACGAGC 2020
Qy	645	leGlnSerLysThrAlaAsp-:::--AspAspCysValIleValAlaLysA 662
Db	2021	TCGACGACATATCCAGGACGATTTGCGGAGTAGACATGGACACAGTCGCTCCAAAC 2080
Qy	662	spGlySerAspTyrAlaSerSerVal-:::--PheAspThrAsnSerGlnGlnLys- 678
Db	2081	AT-:::--ACGCACTATTCACATTTGTCGCATCCCAAGCTCTTCTCTCAAGCCCGAG 2137
Qy	679	----SerLeuAlaSerGlnSerThrGlnLysGlnLeuGlnGlyHisLeuAlaLeuThrT 697
Db	2138	TCGCCAGTCGCTCCACATCAGTCGATCTCGATCTCGAGCAGAA- 2184
Qy	697	hrGlnGlnSerProHisProGlnAsnPheGlnSerThrGlnGlnGlnGlnThrHisLeuA 717
Db	2185	--CAGGAGAATGTGTCAAACTCTCTCCAGGCGGACGACGACCAAGTGGC- 2235
Qy	717	rgMetGluGluMetValThrIleAlaAlaSerSerProLeuPheSerHisHis- 734
Db	2236	-----GCCGCTGCCACCTCAACCTTCGGACAACATTCGCTAA 2272
Qy	735	--AspAspGlnTyrIleAlaGluAlaProThrGluHisTrpGlyArgLysAspAlaLysL 754
Db	2273	GATCCCGGGAATCACTCATCTATTCTCCA-----CACTTATCA--GTGTCACTCGATA 2323
Qy	754	ysLeuThrTrpGluGlnPheLysAlaThrThrArg-----AsnSerProAlaAlaThrC 772
Db	2324	AGACACATGTCNTGCACTACACACTAGTCGACGACCTTCTTCACAAAACCAAGCT 2383
Qy	772	ysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisValMetG 792
Db	2384	ATTCAGGCCAATTT-----CAATCACTTCAT--CGTAAATGCCACCTTCAAG 2428
Qy	792	lysSerSerAsnTyrAlaSerArgGlnProValIleAlaProLeuAspArgTyrAlaG 812
Db	2429	AGTTCATATCCCGGAGCACAGAAATGGGGCTCTCTTGAGCCCG- 2472
Qy	812	luArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMetGluA 832
Db	2473	-----AGCGGGTCGGAACTCGATGTCCAAATATGATT 2506
Qy	832	laSerLysLeuCysAspArgArgAsnAlaGly-----GlnValValLeuTyrProLysG 850
Db	2507	CTTCAGGATCCTACTCGCGCGGTTCGCGAGGTGGAAGCTCTACTGGTATCTATGGAGAG- 2565
Qy	850	luserMetProAlaThrHisLeuArgMetMetAspProSerThrLeuAlaSerPheP 870
Db	2566	-----ACGTTCCAACTCCACGACTATCCGATGAAAAATCCCCGACAT- 2610
Qy	870	roAsnTyrGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHisAsnSerGlnTyrA 890
Db	2611	-----TCTGCCAAAAGTGAGATGGATCCCAACTA- 2640
Qy	890	lahisAsnGlnTyrLysGlySerThrSerThrSerTyrGlySerAsnLeuAsnGlyLysI 910
Db	2641	-----TCATCGCTAGCACGACACATATGATCT--CTCAATGACGAAGT 2683
Qy	910	leProLeuThrPheGluAspLeuSerArgHisGlnLeuHisAspLeuHisArgProLeu- 929



```

Db      2684 AC-----|||...:||||:|||||
Qy      930 -----ArgProHisProArgValGlyValLeuG 939
Db      2720 AGTGTTACAAGAACACTGTCGACTCACTAACCAAGAACAGGAGAACTATGAGCAATGT 2779
Qy      939 lySerLeuLeuGlnLysGluIleAlaSerTrpSerGluAsnCysGlyThrGlnSer----- 957
Db      2780 TTGATCTTTTGGACAAAGCTTAGAAAACCTCACTCAACACATATGATGATCCAACTTGA 2839
Qy      958 -----GlyTyrLysLeuGlyValSerThrG 966
Db      2840 AGCTGGAAGAGCGAATACGATTCAGGAGGACATGCTCATTTGAGGGATATTAGCAATC 2899
Qy      966 lyIleThrSerHisGlnMetAsnArgLysGlu----- 976
Db      2900 ATCTTGATCCCACTGCTGATCATGCTAACGAAGGCGCTGTGAGCTTCTTGGTCAACCAT 2959
Qy      977 -----HisPheGluAlaLeuAsnSerGlyMetPheSerAlaLysT 990
Db      2960 CTCTGGAATCATGTCATCCCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 3019
Qy      990 rpAsnAlaLeuGlnLysSerValSerSerAlaAspPheLeuSerAlaAlaAsn 1010
Db      3020 AGGAGAGATCAGCTTGAGCTCGTTGGCAAGAAC----- 3054
Qy      1010 erIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHisProLeuAspArgPheV 1030
Db      3055 -----AAGAAGAGCTGGATCGCTCCTCA-----CTCTCCAAGTTCA 3091
Qy      1030 alArgGlnAspIleCysIleThrAsnLysAsn-----ProAlaAspPheThrThrIle 1048
Db      3092 CCAAGAG-----AAGAACAGAACTACGACGAGCACATATGCCATCAATTT 3139
Qy      1048 erAsnAspAsnGluTyrMetAsp 1055
Db      3140 CCGATCTCAAGGAAGCTCTTGAC 3162

RESULT 15
US-09-845-917A-55
; Sequence 55, Application US/09845917A
; Publication No. US20030167538A1
; GENERAL INFORMATION:
; APPLICANT: Bogaert, Thierry
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/845,917A
; CURRENT FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 55
; LENGTH: 4749
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-845-917A-55

Alignment Scores:
Pred. No.: 1,72e-09 Length: 4749
Score: 214.00 Matches: 4749
Percent Similarity: 34.68% Conservative: 170
Best Local Similarity: 20.40% Mismatches: 414
Query Match: 3.87% Indels: 365
DB: 10 Gaps: 56

US-09-721-114-2 (1-1057) x US-09-845-917A-55 (1-4749)
Qy      14 ValGlyThrAsnCysMet-----LeuAlaArgGlyThrGlyAla 27
      :|||:|||||:|||||:|||||

```

```

Db      247 CTGGTCTCGACTGCTCGAAACCTCACCAAAACCGATATTCGACGCGGAAACTTGGTGCA 306
Qy      28 ValalaProValLeuGluLeuThrAlaThrProArgGlnAspAlaAlaGluAlaGly 47
      :|||:|||||:|||||:|||||
Db      307 GTTCTCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 345
Qy      48 ValaspGluProalaglnHisGlnCysGluHisPheSerIleArgGlyTyrValalaLeu 67
      :|||:|||||:|||||:|||||
Db      345 ----- 345
Qy      68 LeuGlnLysAspProLysPheCysSerLeuSerArgIlePheHisAspGlnLysLys 87
      :|||:|||||:|||||:|||||
Db      346 -----AGCTTTCGGCACTGAAAGAAAGATCAGAGAAA 378
Qy      88 CysAspGluHisLysAlaSerSerProPheSerValalaLysPheArgArgTyrAsp 107
      :|||:|||||:|||||:|||||
Db      379 TTGAGCAACTACCCACATCATTCATTCGACCGGTTTCTAAATTACCTCGCCACGT 438
Qy      108 CysSerLysCysLeuAspLysLeuLysThrSerAspAsnGlyThrAlaProArgThrLeu 127
      :|||:|||||:|||||:|||||
Db      439 GTCGCCAGCTCAGCAACCGCTTCAGCAACTAACCCAAATTCCAACTTTCACAAATGTCA 498
Qy      128 ProAlaLys---GlnAsnGlyThrSerAspGlyCysSerIleThrPheValaArgSerThr 146
      :|||:|||||:|||||:|||||
Db      499 ACATCCAGCTTCAGACTCCACAGTCAAGAATATCGAAATTCATTCATCAAGATTGGT 558
Qy      147 PheValProAlaSerValGlySerGlnLysValSerProSerThrGlnSerSerGlnGly 166
      :|||:|||||:|||||:|||||
Db      559 ATCAGCCAAAGAGCTTCGACTTAACCCACCTCATCATCAACCTTCATCAAAATAT 618
Qy      167 LysAlaAlaAspArgSerThrLeuProLysSerValGlnGluGlyAsnAsp----- 183
      :|||:|||||:|||||:|||||
Db      619 ACAATATTCATTCCGT-----CGTCCGAGCGTTTCGAGTGGCAATAATAATTGTTGGC 669
Qy      184 -----SerLysCysAsnAlaProSerGlyLysAsnGlyAlaAlaGluAla 198
      :|||:|||||:|||||:|||||
Db      670 TCGACGATATTCACATCTCGAAGAGCTAGATATCATCAACGTCACACTCTATTTCG 729
Qy      199 AsnThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValalaAla 218
      :|||:|||||:|||||:|||||
Db      730 AATCTAAACCGACTACTCTCCCAACTCCAAAACCTCTAGACCAACAAACCCAGTAGTT 789
Qy      219 AsnValSerGluAspAsnThrSerValAspValGlyValalaLeuProGluValProGlnIle 238
      :|||:|||||:|||||:|||||
Db      790 CGTGTTCGT-----ACACTACAAAATCGGAAGC----- 819
Qy      239 ThrTrpHisIleGluValAsnGlyAlaAspGlnProProSerThrProLysLeuSerGlu 258
      :|||:|||||:|||||:|||||
Db      820 -----TCAAAGCTAGCGCTCCGAAAGCGTGCAGCAGCCCAAACTTGTCTCT 867
Qy      259 ValVal-----LeuLysArgAsnGluAspGluAsnGlyLysThrGluThrLeu 275
      :|||:|||||:|||||:|||||
Db      868 GTGAAGACTATTTCGAGCAAAACAGAGCCCGATACACGCGGTGCTGCTGCGAATG 927
Qy      276 ValalaGluGlnCysAsnLeuThrLysAspPro-----AsnProMetSer 290
      :|||:|||||:|||||:|||||
Db      928 CTGAATTAAGTTATTTCAGTACAAAACCCATCTTCTCATCGAATAGCCCAACCT 987
Qy      291 GlyLysGluArgAspGlnValalaGluGlnCysAsnLeuThrLysAspProLysProVal 310
      :|||:|||||:|||||:|||||
Db      988 ACGAGAAAGCGCGCGTCTCAACAACTTGTGCAAAATCGCTGCCCGCAGTG 1047
Qy      311 ---SerGlyGlnLysCysGluGlnIleCysAsnGluProCysGluGluValalaLeuLys 329
      :|||:|||||:|||||:|||||
Db      1048 AAAAGTGGCTGAG-----CCGCGGACCATAGTGGGAGT 1086
Qy      330 ArgSerSerLysSerLys-----ArgLysThrAspLysLys 341
      :|||:|||||:|||||:|||||
Db      1087 GCCACGCTATGTCGAAGCTTTGTACGCCAAAAGTTTCTTACCGTAAACGAGCGCCCA 1146
Qy      342 LeuMetLysLysGlnGlnHisSerLysLysArgThrAlaGlnAlaAspValSerAspAla 361
      :|||:|||||:|||||:|||||
Db      1147 ATCATATCTCAACA-----GACTCG 1167

```



Search completed: May 17, 2004, 13:07:08  
Job time : 2781 secs

Search completed: May 17, 2004, 13:07:08  
Job time : 2781 secs

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	820	14.8	1025	28	BZ795917	BZ795917 PUFHM45TD
C 2	814.5	14.7	888	28	CG355521	CG355521 PUFPL93TB
C 3	716.5	13.0	885	29	CG091207	CG091207 PUFK24TD
C 4	606.5	11.0	787	29	CG345663	CG345663 OGMLW77TH
C 5	567.5	10.3	999	28	BZ795914	BZ795914 PUFHM45TB
C 6	518.5	9.4	766	28	CG603916	CG603916 OGWMH55TV
C 7	491.5	8.9	641	28	BZ411511	BZ411511 OGACC43TM
C 8	465.5	8.4	849	28	BZ411504	BZ411504 OGACC43TC
C 9	414	7.5	523	12	BZ246359	BZ246359 BJ246359
C 10	409	7.4	469	29	AG213697	AG213697 Orvza sat
C 11	401.5	7.3	556	13	BU974305	BU974305 HB27110F
C 12	389	7.0	593	13	BQ244810	BQ244810 TabE15035B
C 13	371.5	6.7	653	28	BZ311518	BZ311518 ic68h08.b
C 14	349	6.3	822	29	CG365016	CG365016 OGHAD70TV
C 15	330	6.0	995	29	CG451574	CG451574 OG8AD34TV
C 16	320	5.8	469	9	AL818058	AL818058 AL18058
C 17	310	5.6	734	28	BZ326387	BZ326387 ic68h08.g
C 18	302.5	5.5	753	28	CG160109	CG160109 ig05g11.b
C 19	289	5.2	792	28	BZ422300	BZ422300 id5a11.b
C 20	287	5.2	491	29	AG209720	AG209720 Oryza sat
C 21	286	5.2	905	29	CG365008	CG365008 OGHAD70TH
C 22	275	5.0	970	29	CG032005	CG032005 PUF2A25TD
C 23	270.5	4.9	782	29	CG824592	CG824592 ZMHBb016
C 24	266.5	4.8	903	28	BZ704364	BZ704364 PUFLE78TD
C 25	259.5	4.7	635	28	BH880296	BH880296 ht52d11.b
C 26	251.5	4.6	688	28	BZ776889	BZ776889 ii04h10.g
C 27	245.5	4.4	600	29	CG603904	CG603904 OGWMH55TH
C 28	241.5	4.4	703	28	BZ776888	BZ776888 ii04h10.b
C 29	241	4.4	870	29	CG091452	CG091452 PUF8F81TB
C 30	240	4.3	701	28	BZ374744	BZ374744 ie31d12.g
C 31	228.5	4.1	590	28	CG160110	CG160110 ig05g11.g
C 32	219	4.0	5400	29	AY418797	AY418797 Homo sapi
C 33	215	3.9	842	28	BZ705032	BZ705032 PUFCS90TD
C 34	214.5	3.9	587	28	BZ359680	BZ359680 id80b06.b
C 35	214	3.9	628	29	CG744988	CG744988 ZMHBb012
C 36	203	3.7	2030	29	CG756375	CG756375 P051-3-F0
C 37	202	3.7	1879	12	BG168882	BG168882 602120215
C 38	200	3.6	664	13	BQ242025	BQ242025 TabE15035B
C 39	200	3.6	1552	12	BI822314	BI822314 603040353
C 40	196.5	3.6	562	28	BZ652753	BZ652753 OGAM082TC
C 41	195	3.5	4773	29	AY411030	AY411030 Homo sapi
C 42	194.5	3.5	793	28	BZ423675	BZ423675 id52a11.g
C 43	194.5	3.5	1894	12	BG113113	BG113113 602283772
C 44	193.5	3.5	183	12	BZ525288	BZ525288 BJ252288
C 45	193	3.5	3877	29	AY414473	AY414473 Mus muscu

# ALIGNMENTS

RESULT 1	BZ795917/c	1025 bp	DNA	linear	GSS 17-MAR-2003
LOCUS	PUFHM45TD	ZM_0.6_1.0_KB	Zea mays	genomic clone	ZM8T2A20H17,
DEFINITION	genomic survey sequence.				
ACCESSION	BZ795917				
VERSION	BZ795917.1				
KEYWORDS	GSS.				
SOURCE	Zea mays				
ORGANISM	Zea mays				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD				
	clade; Panicoideae; Andropogoneae; Zea.				

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 17, 2004, 04:35:29 ; Search time 6194 Seconds  
(without alignments)  
5095.952 Million cell updates/sec

Title: US-09-721-114-2  
Perfect score: 5526  
Sequence: 1 MEIVAVDQEGARVVGTCML.....NRKNPADFTTISNDNEYMDYR 1057

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Zgapop 6.0, Zgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US09721114/runat\_14052004\_092252\_4125/app query.fasta\_1.1223  
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09721114 @CGN 1.14565 @runat\_14052004\_092252\_4125 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estm:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_dg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*

REFERENCE  
AUTHORS  
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J., and  
Bennetzen, J.

TITLE  
JOURNAL  
COMMENT  
Maize Genomics Consortium  
Unpublished (2003)  
Other GSSs: PUFH45TB  
Contact: Cathy Whitelaw

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

FEATURES  
source

Location/Qualifiers

1..1025

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZM061017"

/clone\_lib="ZM 0.6-1.0 kb"

/note="vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high"

CoT selected genomic DNA library"

ORIGIN

Alignment Scores:

Pred. No.: 5,98e-56 Length: 1025

Score: 820.00 Matches: 179

Percent Similarity: 64.60% Conservative: 40

Best Local Similarity: 52.80% Mismatches: 94

Query Match: 14.84% Indels: 26

DB: 28 Gaps: 9

US-09-721-114-2 (1-1057) x BZ795917 (1-1025)

Qy 678 lvsSerLeuAlaSerGlnSerThrGln---LysGluLeuGlnGlyHis---LeuAlaLeu 695  
Db 982 AAATGTTTCAGCCTCCAGAACATCAAAATGCAAAATGCGCTTCAGGTTCATGTGTACCCATC 923  
Qy 696 ThrThrGlnGluSerProHisProGlnAsn-----PheGlnSer 708  
Db 922 AAACGGTCTTCATGTTGATCCCTCAAAATTAACCTGCTTCATGACATTTGGAGTGT 863  
Qy 709 ThrGlnGlnGlnThrHisLeuArgMetGluGluMetValThrLeuAlaAlaSerSer 728  
Db 862 ACCAAGAACACCCAGACACCAATTCACAGGGATGGGGTCCACCATTCATGCTACCTCA 803  
Qy 729 ProLeuPheSerHisHisAspAspGlnTyrlleAlaGluAlaProThrGluHisTTPGly 748  
Db 802 CCTATGTTTCACATCAT-----CAACATATGCTGAAGTGTCTGCTCGAGCTGGAGG 749  
Qy 749 ArgLysAspAlaLysLysLeuThrTTPGluGlnPheLysAlaThrThrArgAsnSerPro 768  
Db 748 AATAACGGGGAAGAGAGTTAATGTGGGATTCCTTCAAGACAGCTTCAAGAAATTCACCA 699  
Qy 769 AlaAlaThrCysGlyAlaGlnPheArgProGlylleGlnAlaValAspLeuThrSerThr 788  
Db 688 ACATCGCATATGTTTTCATTCGGAAACAGGCTTCAGAGATTCATTCCTCCAGT 629  
Qy 789 HisValMetGlySerSerSerAsnTyrlleAlaSerArgGlnProValilleAlaProLeuAsp 808  
Db 628 CATGCTTATGAGCTTCTAGTAACATATGACCTCACCAGCCAGTAATTTGGTGGCTGTAGAT 569  
Qy 809 ArgTyrlleAlaGluAlaValAsnGlnValHisAlaArgAsnPheProSerThrilleAla 828  
Db 568 CAG-----CCGAGAGTGTTCACAGCAACATCG 539  
Qy 829 ThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValValLeuTyrlPro 848  
Db 538 ACTATGAGGCTGTAGTGTATGATCGAGGATTCCTGACAGCTCAGGCTTACCCA 479

Qy 849 lvsGluSerMetProAlaThrHisLeuLeuArgMetMetAspProSerThrLeuAlaSer 868  
Db 478 AAGAGCCCTATGCCCTGCAACACATTTCTTGAGTTGATGGACTCATCAACACTCCAGAC 419  
Qy 869 PheProAsnTyrlleThrSerSerArgAsnGlnMetGluSerGlnLeuHis-----Asn 886  
Db 418 TTCACAACTATCAAGAGCTAACAGAGCCAGAGCCAGAGTGAACATCAACACAACTCTGGAT 359  
Qy 887 SerGlnTyrlleAla---HisAsnGlnTyrlleGlySerThrSerThrSerTyrlleGlySerAsn 905  
Db 358 TCACATATACACAGCATGATCATGATATATGAGTCTCTGAGCAGCATCATATGAGAGACAC 299  
Qy 906 LeuAsnGlyLysilleProLeuThrPheGluAspLeuSerArgHisGlnLeu---HisAsp 924  
Db 298 ATAACCTGGAAGAGTTCACCCGACACTGCAAGACTTATCTCGGCGTCAGGTCCAGCAAAAC 239  
Qy 925 LeuHisArgProLeuArgProHisProArgValGlyValLeuGlySerLeuGlnLys 944  
Db 238 CTGCACAGGCTTTAGCCCTCATCTCGGGTGGGCGCTCTGGTTCGTTGCTGTCAGCAG 179  
Qy 945 GluileAlaAsnTrpSerGluAsnCysGly---ThrGlnSerGlyTyrlleGlyVal 963  
Db 178 GATATCGCAACTGTTATGAAACTCGGGCGCCGACAGCTCTGGATACAGGCTGGCGTT 119  
Qy 964 SerThrGlylleThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSer 983  
Db 118 FCTAAGGGACCGGCATCGCTTGATATGACACACATGGAACATACGAGACCTTGAACCTCA 59  
Qy 984 GlyMetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerAla 1002  
Db 58 GGAGTGTTCACAGCAGGATGGATGCGCTTCAATGGGTCTCTGTAGCTCTCTTGGC 2

RESULT 2

LOCUS

CC355521

DEFINITION

genomic survey sequence.

ACCESION

CC355521

VERSION

CC355521.1

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,

Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J., and

Bennetzen, J.

Maize Genomics Consortium

Unpublished (2003)

Other\_GSSs: PUFH45TB

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..888

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZM061017"

/clone\_lib="ZM 0.6-1.0 kb"

/note="vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high"

CoT selected genomic DNA library"

ORIGIN

Alignment Scores:

Pred. No.: 1.35e-55 Length: 888  
Score: 814.50 Matches: 169  
Percent Similarity: 69.83% Conservative: 37  
Best Local Similarity: 57.29% Mismatches: 72  
Query Match: 14.74% Indels: 17  
DB: 28 Gaps: 6

US-09-721-114-2 (1-1057) : CC355521 (1-888)

```
Qy 767 SerProAlaAlaThrCysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThr 786
Db 2 TCACCAACATCTGATATGGTTCCTTCAATTCGGAACAGGCTTCAAGAAGTTCATTCATCT 61

Qy 787 SerThrHisValMetGlySerSerSerSerSerSerSerSerSerSerSerSerSerSer 806
Db 62 CCGAGTCATGCTTATGGAGCTTCTAGTAATCTATGAGCTCACCAGCCAGTAATGTGGCT 121

Qy 807 LeuAspArgTyAlaGluAlaValAsnGlnValHisAlaArgAsnPheProSerThr 826
Db 122 ATAGATCAG-----CCGAGAAGTGTTCACAGCACA 151

Qy 827 IleAlaThrMetGluAlaSerLysLeuCysAspArgArgAsnAlaGlyGlnValValLeu 846
Db 152 ACATCGACTATGGAGCTGGTGGTGTGATGATGAGGATTCCTGGACAGTCAGGCCTC 211

Qy 847 TyrProLysGluSerMetProAlaThrHisLeuLeuArgMetMetMetMetMetMetMet 866
Db 212 TACCCAAAGAGCCCTATGCTGCAACACATCTCTGACATGTGAGCTCATCAACAGCT 271

Qy 867 AlaSerPheProAsnTyGlyThrSerSerArgAsnGlnMetGluSerGlnLeuHis---- 885
Db 272 CCAGACTTTCACAAACTATCAAGAGCTAACAGGCCAGATGGAAACATCAACAACT 331

Qy 886 ---AsnSerGlnTyAla---HisAsnGlnTyLysGlySerThrSerThrSerTyGly 903
Db 332 CTGGATTCACATATACACAGCATGATCAGTATATGAGTCTACTGACGACATCATATGA 391

Qy 904 SerAsnLeuAsnGlyIlePheLeuThrPheGluAspLeuSerArgHisGlnLeu--- 922
Db 392 AGACACATAAATCGGAAGAGTTCACCGCACTGCAAGACTTATCTCGCGCGTCAGTCCAG 451

Qy 923 HisAspLeuHisArgProLeuArgProHisProArgValGlyValLeuGlySerLeuLeu 942
Db 452 CAAACCTTCACAGGCGCTTACGCCCTCTCTCGGGTGGGGCTGTGGTGGTGGTGGTGGT 511

Qy 943 GlnLysGluIleAlaAsnTrpSerGluAsnCysGly---ThrGlnSerGlyTyTrpLysLeu 961
Db 512 CAGCAGGATATCGAAACTGGTATGAAACTCGCGGCCCGCACAGCTCTGGATACAGGCTG 571

Qy 962 GlyValSerThrGlyIleThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeu 981
Db 572 GCGGTTCCTAAAGGACCGCATCGCTGATGATGAACACACATGGAACATACGAGACCTTG 631

Qy 982 AsnSerGlyMetPheSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerSer 1001
Db 632 AACTCAGAGTGTTCACAGCAGATGGAATGCCCTTCAATTTGGTTCCTGTAGCTCTCTT 691

Qy 1002 AlaAsp-----PheLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLys 1019
Db 692 GCCGATCCAGATACCGGTTCGCAAGGTATGGTACAGGTACAGCTTGTACAGGTGGCAAT 751

Qy 1020 GlyValMetValHisProLeuAsnArgPheValArgGlnAspIleCysIleThrAsnLys 1039
Db 752 GGGAGAACGTTTACCCCGTTGGATAGCTTGTGGCAAGGATATCTGTGTGATCAACAGA 811

Qy 1040 AsnProAlaAspPheThrThrIleSerAsnAspAsnGluTyMet 1054
Db 812 AACCCAGCTGATTTTACTGTAATAGTGACATGAATGAGTATATG 856
```

RESULT 3

CG091207

LOCUS

DEFINITION

CG091207 885 bp DNA linear GSS 20-AUG-2003  
PUJFK24TD ZM\_0.6\_1.0\_KB Zea mays genomic clone ZM56C23,

# genomic survey sequence.

CG091207 GI:33973501  
KEYWORDS  
SOURCE

ORGANISM  
Zea mays

Zeal mays  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

### AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and  
Bennetzen, J.

## TITLE

### JOURNAL

### COMMENT

Maize Genomics Consortium  
Unpublished (2003)  
Other GSSs: PUJFK24TB  
Contact: Cathy Whitelaw

## TISSUE

7912 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

## FEATURES

### source

1..885  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZM56C23"  
/clone\_lib="ZM\_0.6\_1.0\_KB"  
/notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
COT selected genomic DNA library"

## ORIGIN

### Alignment Scores:

Pred. No.: 1.18e-47 Length: 885  
Score: 716.50 Matches: 160  
Percent Similarity: 60.97% Conservative: 29  
Best Local Similarity: 51.61% Mismatches: 88  
Query Match: 12.97% Indels: 33  
DB: 29 Gaps: 9

US-09-721-114-2 (1-1057) x CG091207 (1-885)

```
Qy 753 LysLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsn-----SerProAlaAla 770
Db 27 AAGAAGCTAATGTGGATTCCTTTCAACACAGCTTCAAGGAATCCACCATCACCACACCG 86

Qy 771 ThrCysGlyAlaGlnPheArgProGlyIleGlnAlaValAspLeuThrSerThrHisVal 790
Db 87 TCATATGCTTTTCAATTCGGAACAGGCTTCAAGAAAGTTGATTCAGCTCCCATTCATGCA 146

Qy 791 MetGlySerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 810
Db 147 -----TATGGAGCTCACCAGCCAGTAAATTTGGCGGTAGATCAG--- 185

Qy 811 AlaGluArgAlaValAsnGlnValHisAlaArgAsnPheProSerThrIleAlaThrMet 830
Db 186 -----CCGAGAAGTGTTCACAGCAGCGGTGGCTATG 218

Qy 831 GluAlaSerLysLeuCysAspArgArgAsnAlaGlnValValLeuTyProLysGlu 850
Db 219 GAGCTGTAGGCTGTACGATCAAGAGTTCCTGGACGCGCCGCTTTACCCAAAGAG 278

Qy 851 SerMetProAlaThrHisLeuLeuArgMetMetMetMetMetMetMetMetMetMetMet 870
Db 279 CCCATGCTGCAACACATCTCTGAGGCTGATGAGTCTCATCAACAGCTCCAGGCTTACA 338

Qy 871 AsnTyGlyThrSerSerArgAsnGlnMetGlu-----SerGlnLeuHisAsnSerGln 888
Db 339 AACTACCAAGAGCTAATAGCGGCGAGATGAGCTTCAACAGCAAGCTCTCGGCTCGAG 398
```

methylation filtered genomic DNA library"

[illegible]

```
/clone_lib="ZM 0.7 1.5_KB"
```

QY	702	HisProGlnAanPhe	766	bp	DNA	linear	GSS 18-JUN-2003
LOCUS	CG603916		766	bp	DNA	linear	GSS 18-JUN-2003
DEFINITION	CG603916		766	bp	DNA	linear	GSS 18-JUN-2003
ACCESSION	CG603916		766	bp	DNA	linear	GSS 18-JUN-2003
VERSION	CG603916.1		766	bp	DNA	linear	GSS 18-JUN-2003
KEYWORDS	GSS		766	bp	DNA	linear	GSS 18-JUN-2003
SOURCE	Zea mays		766	bp	DNA	linear	GSS 18-JUN-2003
ORGANISM	Zea mays		766	bp	DNA	linear	GSS 18-JUN-2003
REFERENCE			766	bp	DNA	linear	GSS 18-JUN-2003
AUTHORS			766	bp	DNA	linear	GSS 18-JUN-2003
TITLE			766	bp	DNA	linear	GSS 18-JUN-2003
JOURNAL			766	bp	DNA	linear	GSS 18-JUN-2003
COMMENT			766	bp	DNA	linear	GSS 18-JUN-2003
FEATURES			766	bp	DNA	linear	GSS 18-JUN-2003
source			766	bp	DNA	linear	GSS 18-JUN-2003
ORIGIN			766	bp	DNA	linear	GSS 18-JUN-2003
Alignment Scores:			766	bp	DNA	linear	GSS 18-JUN-2003
Pred. No.:			766	bp	DNA	linear	GSS 18-JUN-2003
Score:			766	bp	DNA	linear	GSS 18-JUN-2003
Percent Similarity:			766	bp	DNA	linear	GSS 18-JUN-2003
Best Local Similarity:			766	bp	DNA	linear	GSS 18-JUN-2003
Query Match:			766	bp	DNA	linear	GSS 18-JUN-2003
DB:			766	bp	DNA	linear	GSS 18-JUN-2003
US-09-721-114-2			766	bp	DNA	linear	GSS 18-JUN-2003
QY	626	LysAsnGlnHisGluArgGlnLeuMetThrGluThrAspCysSerAspIleAsnArgIle	645				
DB	27	CGAAATACGATGAGAGACAGTTCATGATCAGACCAATCTTTGGAAAACCTGCATAC	86				
QY	646	GlnSerLysThrThrAlaAspAspCysValIleValAlaLysAspGlySerAsp	665				
DB	87	CAACCCAG--GTAGCTCAAGTATGCTGCTGATTCAGCAGAGGTTTCCAC	143				
QY	666	TyrAlaSerSerValPheAspThrAsnSerGlnGlnLysSerLeuAlaSerGlnSerThr	685				
DB	144	AATGCATCAATAGAGTTCACACTAATTC--CAAAAATCTTTGGCATCGGAAGTAA	200				
QY	686	GlnLysGluLeuGlnGlyHisLeuAla--	694				
DB	201	CAGAAGTCTTACAGATCAGATCACCAGTACAGAGGCTGCCAATGTGAATCCGAG	260				
QY	695	---LeuThrGlnGlnSer---	700				
DB	261	GATTTACATCTACAGAGTTCATGATGATGATGATGATGATGATGATGATGATGAT	320				
QY	701	---	701				
DB	321	CCAAATGGTCAGCTCCAGATCAAAATGCAAAATTCGCTTCAGGTTTCATGTGTTACCC	380				





Db	785	GTACACACGATGAGCATTTACAAATCGGTCCACGAGCACCGTCATATATGGAAGCCAA-----	731
Qy	907	nGlyLysIleProLeuThrPheGluAspLeuSer---AtqHisGlnLeuHisAspLeuHi	926
Db	730	-----CCGCTGACACTGGAAGACTTGTCTCGCGGTCAATTCAGCAAGACTTGGCG	681
Qy	926	sArgProLeuArgProHisProArgValGlyValLeuGlySerLeuLeuGlnLysGluIl	946
Db	680	CAGCGCTTTACGCGCTTCAACCTCGTGTGGGTGTCTCGGTTCATTGCTGAGGAGAGAT	621
Qy	946	eAlaAsnTrpSerGluAsnGlyThrGlnSerGlyTyrLysLeuGlyValSerThrGl	966
Db	620	CGCAAACTGTCTGGGACTCGCGCGCGAGTCTGGGTACAGACTAGTGTGATGTTGTAAGG	561
Qy	966	yIleThrSerHisGlnMetAsnArgLysGluHisPheGluAlaLeuAsnSerGlyMetPh	986
Db	560	GACAAGCGCGCTGATGTCACACAGAGCAGAGAACTACGAGACCTTGAGC-----	512
Qy	986	eSerAlaLysTrpAsnAlaLeuGlnLeuGlySerValSerSerAla-----AspPh	1004
Db	511	-TCGGCAGGATGAAACACCTCGAGTGGGTCTGTAGCTCTGCTGCCAATCTGGAGTA	453
Qy	1004	eLeuSerAlaArgAsnSerIleAlaGlnSerTrpThrArgGlyLysGlyLysMetValHi	1024
Db	452	CCGGTTTCCATGATGTCGTCAGCGGTGAGCCTTCGACCAGTGGCATGGGAGGACCATTC	393
Qy	1024	sProLeuAspArgPheValArgGlnAspIleCysIleThrAsnLysAsnProAlaAspPh	1044
Db	392	TCGGTGGGTAACCTCGTCAGGAAGGATATCTGTGACTAAACAGAAACCCAGCCGATTT	333
Qy	1044	eThrThrIleSerAsnAspAsnGluTyrMet	1054
Db	332	CATGTAAATAGTCAGACAGACGATGATG	302
RESULT 9			
LOCUS	BJ246359	523 bp mRNA linear EST 05-APR-2002	
DEFINITION	BJ246359 Y. Ogiwara unpublished cDNA library, Wh_f Triticum aestivum cDNA clone whf22j14 5', mRNA sequence.		
ACCESSION	BJ246359		
VERSION	BJ246359.1	GI:20058037	
KEYWORDS	EST.		
SOURCE	Triticum aestivum (bread wheat)		
ORGANISM	Triticum aestivum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.		
AUTHORS	1 (bases 1 to 523)		
TITLE	Ogiwara,Y. and Murai,K.		
JOURNAL	Expressed genes in Triticum aestivum		
COMMENT	Unpublished (2002) Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp.		
FEATURES	Location/Qualifiers		
source	1..523		
	/organism="Triticum aestivum"		
	/mol_type="mRNA"		
	/cultivar="Chinese Spring"		
	/db_xref="taxon:4565"		
	/clone="whf22j14"		
	/tissue_type="spike at flowering date"		
	/dev_stage="Feekes' scale 10.5.1"		
	/clone_lib="Y. Ogiwara unpublished cDNA library, Wh_f"		
ORIGIN			
Alignment Scores:	1.91e-23	Length:	523
Pred. No.:			

Score: 414.00 Matches: 85  
Percent Similarity: 55.16% Conservativity: 16  
Best Local Similarity: 54.84% Mismatches: 54  
Query Match: 7.49% Indels: 0  
DB: 12 Gaps: 0

US-09-721-114-2 (1-1057) x BJ246359 (1-523)

QY 708 SerThrGlnGluGlnGlnThrHisLeuArgMetGluGluMetValThrIleAlaAlaSer 727  
DB 51 TCRACCCAGGAAACACAGGACATCGGGATGGAGATATGTCATATGTCGACCC 110  
QY 728 SerProLeuPheSerHisHisAspGlnTrpIleAlaGluAlaProThrGluHisTrp 747  
DB 111 TCACCACTATTATCACCAACAGACAGCAGCTCTCTGCTGAAGCAACACCTGAGCGGTGG 170  
QY 748 GlyArgLysAspAlaLysLeuThrTrpGluGlnPheLysAlaThrThrArgAsnSer 767  
DB 171 AGCCATATGGGAGCAAGAAGTCGATGTGGAACTTTCAAGGCACCTTCACAGGGATTTA 230  
QY 768 ProAlaAlaThrCysGlyAlaGlnPheArgProGlyIleGlnAlaValAlaAspLeuThrSer 787  
DB 231 TCACCAACACAGAGTGTGCTGCTCAGTTTCAGACCTAGCATTGCACAGCGTTGATTAACTTAT 290  
QY 788 ThrHisValMetGlySerSerSerAsnTrpAlaSerArgGlnProValIleAlaProLeu 807  
DB 291 ACTGATGGGGGAGTAACTGTTATTATCCCACTCGCCAGCAGTAATTTGCACACTT 350  
QY 808 AspArgTyraGluAlaGluAlaValAsnGlnValHisAlaArgAsnPheProSerThrIle 827  
DB 351 GACCACTATCAANTAGAGCAGTTAACTCAGTCGAGCAGAGAGTTTTCAGTTTCAGTG 410  
QY 828 AlaThrMetGluAlaSerLysLeuCysAspArgAsnAlaGlyGlnValLeuTyrr 847  
DB 411 TCACCACTGGAAGCTGGTAACTCTGTGTGATGGAGAAATGTTGCACATTCAGGTTTAT 470  
QY 848 ProLysGluSerMetProAlaThrHisLeuLeuArgMetMetAsp 862  
DB 471 CCAGAGAAACCATGCTCTGCTACTCAGCCCTGAGATTGACCGAG 515

RESULT 10  
AG213697  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) DNA, clone: T10466T, 3' flanking sequence of Tos17 insertion in rice strain ND6047, genomic survey sequence.  
ACCESSION  
AG213697.1 GI:32360887  
VERSION  
GSS.  
KEYWORDS  
Oryza sativa (japonica cultivar-group)  
SOURCE  
Oryza sativa (japonica cultivar-group)  
ORGANISM  
Oryza sativa (japonica cultivar-group)  
REFERENCE  
1 Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K. and Hirochika, H.  
Target Site Specificity of the Tos17 Retrotransposon Shows a Preference for Insertion within Genes and against Insertion in Retrotransposon-Rich Regions of the Genome  
Plant Cell 15 (8), 1771-1780 (2003)  
JOURNAL  
MEDLINE  
PUBMED  
12897251  
REFERENCE  
2 (bases 1 to 469)  
Miyao, A., Miyazaki, A., Yamashita, Y. and Hirochika, H.  
Direct Submission  
TITLE  
Submitted (16-APR-2002) Akio Miyao, National Institute of Agrobiological Sciences, Molecular Genetics, 2-1-2, Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: miyao@affrc.go.jp, URL: http://tos.nias.affrc.go.jp/, Tel: 81-298-38-7020, Fax: 81-298-38-7020)  
FEATURES  
Location/Qualifiers  
1. .469

/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/strain="ND6047"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="T10466T"  
/clone\_lib="PCR product directly amplified from rice genomic DNA"  
/note="The 3' end of retrotransposon Tos17 was found immediately upstream of this sequence."  
misc\_feature  
1  
/note="This insertion point is base 1732 in AF326768."

ORIGIN

Alignment Scores:  
Pred. No.: 4,17e-23 Length: 469  
Score: 409.00 Matches: 79  
Percent Similarity: 100.00% Conservativity: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.40% Indels: 0  
DB: 29 Gaps: 0  
US-09-721-114-2 (1-1057) x AG213697 (1-469)  
QY 498 LeuAspThrAsnMetHisLysThrAspValCysGlnHisValSerGluLeuSerThrGln 517  
DB 1 CTGACACAAATATGATTAAGACAGATGCTGTGACGATGATCAGAAATCTCCACACAG 60  
QY 518 ArgCysSerSerLysLysThrAlaGlyLeuSerLysGlyLysThrHisSerAlaAla 537  
DB 61 AGTGCTCATCAAGGGGAAACAGCGGCTTTGAGTAAGGGGAAACACATTCAGCTGCT 120  
QY 538 SerThrLysTyrrGlyGlyGluSerThrArgAsnGlyGlnAsnIleHisValLeuSerAla 557  
DB 121 AGTACCAATATGGTGGTGAAGACCAAGAAATGTCAGAACATCATGTACTCAGCGCA 180  
QY 558 GluAspGlnCysGlnMetGluThrGluAsnSerValLeuSerHisSerAlaLysVal 576  
DB 181 GAAGATCAATGCCAGATGGAAACCGAAACTCTGTTCTGAGTCACTCGGCAAGGTA 237

RESULT 11

BU974305  
LOCUS  
DEFINITION  
HB27110r BC Hordeum vulgare subsp. vulgare cDNA clone HB27110  
5-PRIME mRNA sequence.  
ACCESSION  
BU974305  
VERSION  
BU974305.1 GI:24225098  
KEYWORDS  
EST.  
SOURCE  
Hordeum vulgare subsp. vulgare  
ORGANISM  
Hordeum vulgare subsp. vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae; Triticeae; Hordeum.  
REFERENCE  
1 (bases 1 to 556)  
Radchuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.  
Barley ESTs from developing seeds  
Unpublished (2002)  
CONTACT  
Contact: Stein Nils  
Molecular Markers Group, Department Genbank  
Institute of Plant Genetics and Crop Plant Research (IPK)  
Corrensstr. 3, 06466, Gatersleben, Germany  
Tel: 039482-5522  
Fax: 039482-5595  
Email: stein@ipk-gatersleben.de  
Insert Length: 556 Std Error: 0.00  
Plate: 27 row: 1 column: 10  
Seq primer: M13rev, Location/Qualifiers  
1. .556  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="barke"  
/sub\_species="vulgare"





Class: sheared ends.  
 Location/Qualifiers  
 1. 822  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="ZM0.7.1.5 KB"  
 /notes="Vector: pBCSK-1; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

FEATURES  
 source  
 Alignment Scores:  
 Pred. No.: 6,56e-18 Length: 822  
 Scores: 349.00 Matches: 109  
 Percent Similarity: 47.49% Conservative: 33  
 Best Local Similarity: 36.45% Mismatches: 90  
 Query Match: 6.12% Indels: 68  
 DB: 29 Gaps: 15

US-09-721-114-2 (1-1057) x CG365016 (1-822)

Qy 63 GlyTyrValAlaLeuLeuGlnLysLysAspProLysPheCysSerLeuSerArgIlePhe 82  
 Db 3 GGTATGTTGCTATGCTTCAGAGAGAGATCCAAAACCTTTGCTCT---CCTCATATTTTC 59  
 Qy 83 HisAspGlnLysCysAspGluHisLysAlaSerSerSerProPheSerValAlaLys 102  
 Db 60 CATAGCCAGCCCAATATGATGAGCACCAT---GGTCTGCTCCCATTTTGGTGTCAAG 116  
 Qy 103 PheArgTrpAspCysSerLysCysLeuAspLysLeuLysLysLeuLysThrSerAspAsnGlyThr 122  
 Db 117 TATCGTCGATGGATGCTCAAAATGCTTGAAGAGATGAAAGTTTCAGGCCATCGACCT 176  
 Qy 123 AlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIleThrPhe 142  
 Db 177 ACATCTGGAATGTTTCTATCGAGCAGAGTGAATGAATGATGGTCTCTATTTCAATT 236  
 Qy 143 ValArgSerThrPheValProAlaSerValGlySerGlnLysVal---SerProSerThr 161  
 Db 237 GTTCCG-----ATTTTGCTATATAGGTGTATCCAGAGAGATGTTTATTTGCAACAA 290  
 Qy 162 GlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGlnGlnGly 181  
 Db 291 CATCATCTCAGGGAT-----GATCAATTGACCTTTCAGAACGCCCAAGATGC 344  
 Qy 182 AsnAspSerLysCysAsnAlaPro----- 189  
 Db 345 AAT---TCAAAATGCAGTTCTCTGTCGCAACAAAGCAATTAAGTATTCAGTA 401  
 Qy 190 ----- 192  
 Db 402 GCTGAGAGAAATGTTCTAGAGGACCTCGTTGAAAGAGTGTCCAGCCAGCAGAGTATG 461  
 Qy 193 AsnGlyAlaAlaGlu-----AlaAsnThrAsp 201  
 Db 462 GCTGGGCGATGTTGAATTTCTTAATTAATCTTTCTCTGTATACAAATAAGCTAATAT-GAC 520  
 Qy 202 SerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAlaAlaAsnValSer 221  
 Db 521 AGTATTTCTCAGATTTACAGGCTCTCCCAATATATATTTGATGTTCTCAGCAAAATATCTG 580  
 Qy 222 GluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIleThrTrpHis 241  
 Db 581 ---AATGTTGTTCCCAAGGATCCAGCGATTTTACCTGATGATGTTTCAAACCATATCAACC 637  
 Qy 242 IleGluValAsnGlyAlaAspGlnProProSerThrProLysLeuSerGluValValLeu 261  
 Db 638 ATTGAGAAATATGGCCAAATAATCATGTAG---CCAAACTTCGT-----GTG 685  
 Qy 262 LysArgAsnGluAspGluAsnGlyLysThrGluGluThrLeuValAlaGlnGlnCysAsn 281  
 Db 685

686 ATGCCAATGAAGATGAGAGT-----AACATAGTTCAAGATGTTCTTAAT 730  
 282 LeuThrLysAspProAsnProMetSerGlyLysGluArgAspGlnValAlaGluGlnCys 301  
 731 TTT-----GACCCCAAT-----GAGTCTAATCTTTGTAACCATTA 766  
 302 AsnLeuThrLysAspProLysProValSerGlyGlnLysCysGlnLysCysAsn 320  
 767 TCTTGCCCAAAAGGC---AAACAATATCTGGGCACAAAGAGCAGCCAGCTGTGCAAC 820

RESULT 15  
 CG451574 995 bp DNA linear GSS 17-SRP-2003  
 LOCUS CG8AD34TV ZM 0.7.1.5 KB Zea mays genomic clone ZM58Ma0848F20,  
 DEFINITION genomic survey sequence.  
 ACCESSION CG451574  
 VERSION CG451574.1 GI:34836574  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 995)  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
 Citek, R.W., Numborg, A., Robbins, D. and Lakey, N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 CONTACT: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TP  
 Class: sheared ends.  
 Location/Qualifiers  
 1. 995  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="ZM0.7.1.5 KB"  
 /clone\_lib="ZM 0.7.1.5 KB"  
 /notes="Vector: pBCSK-1; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,95e-16 Length: 995  
 Scores: 330.00 Matches: 102  
 Percent Similarity: 46.44% Conservative: 35  
 Best Local Similarity: 34.58% Mismatches: 72  
 Query Match: 5.97% Indels: 87  
 DB: 29 Gaps: 14

US-09-721-114-2 (1-1057) x CG451574 (1-995)

Qy 61 IleArgGlyTyrValAlaLeuLeuGlnLysLysAspProLysPheCysSerLeuSerArg 80  
 Db 242 GTCAGGGTATGTTGCTATGCTTCAGAGAGAGATCCAAAACCTTGC-----TCTCAT 295  
 Qy 81 IlePheHisAspGlnLysLysCysAspGluHisLysAlaSerSerSerProPheSerVal 100  
 Db 296 ATTTTCCATAACAGCCCAATATGATGAGCACCAT---GATTGCTCCCACTTTTGGTG 352  
 Qy 101 AlaLysPheArgArgTrpAspCysSerLysCysLeuAspLysLeuLysThrSerAspAsn 120  
 Db 353 TCGAGTATCATCGATGGGATGCTCAAGATGCTCTTGATAGTGTGAGATTTTCAGGCCAT 412  
 Qy 121 GlyThrAlaProArgThrLeuProAlaLysGlnAsnGlyThrSerAspGlyCysSerIle 140  
 Db 140

Db 413 AGGCCAACATCGAAAAATGTTTCTATGACGACGAATGAATGAATGATGGCTGCTCTATT 472  
Qy 141 ThrPheValArgSerThrPheValProAlaSerValGlySerGlnLysVal---SerPro 159  
Db 473 TCATGTTGCGG-----ATTGCTCTAATAGTGTGTGATTCGAGGACTGTTTCTTCG 526  
Qy 160 SerThrGlnSerSerGlnGlyLysAsnAlaAspArgSerThrLeuProLysSerValGln 179  
Db 527 ACACAGCAATCATCTCAAGGAAT-----GATCGATTGACCCCTTCAAAGACTGCCCAA 580  
Qy 180 GluGlyAsnAspSerLysCysAsnAlaProSerGlyLys----- 192  
Db 581 AAATGCAAT---TCAAAATGCGAGTTCTCTCGCAACAAGCAATTAATGCGATGAATGTT 637  
Qy 193 -----Asn 193  
Db 638 CCAAGTAGCTGAAGAAATGTTCCAGAGGCACTCGTTGATACGAGGGTTCATCTATAGAA 697  
Qy 194 GlyAlaAlaGlu-----AlaAsn 199  
Db 698 GGTATGGCTGACATGTTGAATCTTAATTACTTTTGTGTGTATACAAGCTAAT 757  
Qy 200 ThrAspSerProMetLysAspLeuGlnGlyProAlaGlnAsnTyrAspValAlaAlaAsn 219  
Db 758 AT-GACAGTATTTCTCAGTTTTCAGGCTCTCCCAATAGTATGATTGTTGTCAGCAAT 816  
Qy 220 ValSerGluAspAsnThrSerValAspValGlyAlaLeuProGluValProGlnIleThr 239  
Db 817 ATCTTG---AATGCTGTCTCGAAGAATGTCAGCGATCTACCTGAT----- 858  
Qy 240 TrpHisIleGluValAsnGlyAlaAspGlnProProSerThrProLysLeuSerGluVal 259  
Db 859 ---GATGTTGAAGAAATGGCACAGATCCACAGTCTA-----AAACCT 903  
Qy 260 ValLeuLysArgAsnGlnAspGluAsnGlyLysThrGluGluThrLeuValAlaGluGln 279  
Db 904 TGTGTGGCGCCAAATGAAGATTAGAATAACATAGCTAATGATGACCTC----- 951  
Qy 280 CysAsnLeuThrLysAspProAsnProMetSerGlyLysGluArgAspGlnValAlaGlu 299  
Db 952 -----AATGAGCCTAAT----- 963  
Qy 300 GlnCysAsnLeuThrLysAspProLysProValSerGlyGlnLys 314  
Db 964 GTATGT-----AAACCAGTATCTGGGCACAAA 990

Search completed: May 17, 2004, 09:59:24  
Job time : 6235 secs